The Utility of Webcam Eye-Tracking: Lessons Learned and Practical Applications in Decision-Making and Digital Human-Computer Interactions

by

Jennifer K. Bertrand

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Faculty of Kinesiology, Sport, and Recreation
University of Alberta

Abstract

In the realm of contemporary research, especially after a global pandemic, the potential of remotely-collected webcam eye-tracking has emerged as a promising yet relatively uncharted avenue. This thesis aims to both uncover and bolster the utility of this method in understanding human behaviors, navigating the intricate trade-offs it presents.

Chapter 1 gives context to the aim of this thesis by providing a panoramic view of eye-tracking, remote and online experimentation, and the unique challenges that they present when combined to form remote webcam eye-tracking.

In Chapter 2, we endeavoured to understand the utility of webcam eye-tracking as a tool to explore the persistence of eye-hand coordination patterns within digitized object interactions. Operating within a fully remote experimental setup, we successfully deployed a digital drag-and-drop cursor movement task with webcam eye-tracking, gathering complete datasets from 51 participants who used their own web-cameras as eye-trackers. Our results demonstrate the usefulness of webcam eye-tracking by capturing robust eye-cursor dynamics, and further reveal novel empirical insights about the adaptability of the visuomotor system in a digital domain.

Chapter 3 pivots to evaluate the practical utility of webcam eye-tracking within a binary choice decision-making domain. Building off of prior cursor-tracking work and established binary choice tasks, we deployed our webcam eye-tracking study to 100 remote, crowdsourced participants. With webcam eye-tracking, we found decision-difficulty effect earlier than what cursor-tracking alone could provide, demonstrating the utility of webcam eye-tracking. In addition, webcam eye-tracking afforded us

the opportunity to capture more nuanced decision-making processes as they relate to gaze-driven information sampling, adding depth to the method's usefulness.

Accompanying these investigations, Chapter 4 serves as a comprehensive, practical guide, crystallizing lessons from the challenging process of applying webcam eye-tracking to our chosen domains. Beyond its value as a collection of invaluable tips, the chapter leverages evidence from the previous studies, providing evidence-based recommendations on data quality, participant engagement, and experimental costs.

This thesis concludes with Chapter 5, providing a reflective perspective when revisiting the contributions made in the previous chapters. Analyzing the method's utility and its empirical implications, my definition of utility evolves to be more broad and holistic. I contemplate the future of eye-tracking technology from various angles, and conclude this thesis with personal reflections on the challenges and rewards of harnessing webcam eye-tracking.

In summary, this thesis showcases the potential of remotely-collected webcam eye-tracking in comprehending human behaviors. From eye-hand coordination to decision-making, and from practical guidance to future implications, the research paints a compelling narrative of webcam eye-tracking's utility, bridging the gap between remote data collection and nuanced insights into human behavior.

Preface

This thesis is an original work by Jennifer K. Bertrand. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name "ACE 2", No. Pro00087329, February 25, 2019.

Chapter 2 of this thesis has been published as: Bertrand, J. K. & Chapman, C. S. (2023). Dynamics of eye-hand coordination are flexibly preserved in eye-cursor coordination during an online, digital, object interaction task. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, April 23–28, 2023, Hamburg, Germany. ACM, New York, NY, USA, 13 pages. doi: 10.1145/3544548.3580866. Data were collected by myself using online participants, with experimental procedures approved by the University of Alberta's Research Ethics Office. I was responsible for the majority of the work across all aspects of the experimental process, with Craig Chapman's consultation and support throughout.

Chapter 3 of this thesis has not yet been published in a peer-reviewed journal, but has been published on a pre-print server as Bertrand, J. K., Zuk, A. A. O., & Chapman, C. S. (2023). Continuous Measures of Decision-Difficulty Captured Remotely: II. Webcam eye-tracking reveals early decision processing. *bioRxiv*, 2023-06. doi:10.1101/2023.06.06.543799. Data were collected by myself using online participants, with experimental procedures approved by the University of Alberta's Research Ethics Office. This study is the second study in a two-part series, where the first study has also been published on a pre-print server as Zuk, A. A. O., Bertrand, J. K., & Chapman, C. S. (2023). Continuous Measures of Decision-Difficulty Captured

Remotely: I. Mouse-tracking sensitivity extends to tablets and smartphones. *bioRxiv*, 2023-06. doi:10.1101/2023.06.06.543796. The over-arching concept and design of the two experiments was the work of Alexandra Ouellette Zuk and Craig Chapman, with the Chapter 3 experiment building off some implementation and programming efforts from the preceding study (as performed by Alexandra Ouellette Zuk and Craig Chapman). I performed additional implementation and data processing for this experiment, and was responsible for the complete data collection process. I was also responsible for the majority of data analysis, visualization, and data curation, with support from Craig Chapman. I was primarily responsible for the manuscript writing, with writing and manuscript composition involvement from Alexandra Ouellette Zuk and Craig Chapman.

Chapter 4 of this thesis has not yet been published. It is a methodological contribution that uses some data from the studies of Chapters 2 and 3, both collected with approval of the University of Alberta's Research Ethics Board (as stated above). Some of the paper's concepts relate to some of the early concept discussions with Caspar Goeke and Holger Finger in 2021-2022 as part of my Research Scientist internship with Scicovery GmbH (Mitacs Accelerate International program). Further conceptualization was performed by myself and Craig Chapman, and I was responsible for the majority of the analysis, visualizations, data curation and writing, with support and consulation throughout from Craig Chapman.

Chapters 1 and 5 are original works composed by Jennifer K. Bertrand, and are previously unpublished.

To my favourite person and partner in life, Luke (and our little pup too).

Acknowledgements

I am humbly grateful for the village of support that has shaped this thesis.

First, I am profoundly grateful to my other half, Luke. It's hard to keep up with the smartest person I know, which may explain why you have only known me as a university student for the past 10+ years of our life together. You have supported me without reservation, in every way, making it possible to build a beautiful life together despite my perpetual student status. You even (finally) let us get the neurotic pandemic puppy of our dreams, Tumi. Your love has made all of life's challenges manageable, and the achievements that much sweeter. I will forever be grateful that I completed this PhD with you by my side. Thank you a million times over.

My family and close friends, both by blood and by heart, have been a constant source of enthusiasm for my academic pursuits. Their love and care have greatly enriched my life outside of academia.

My academic village is vast, and I am privileged to have outstanding mentors and colleagues:

Dr. Craig Chapman, my supervisor for nearly a decade, has been instrumental in shaping my academic and professional path. Your support, encouragement, and unwavering passion for science have been invaluable. Our data meetings and your contagious excitement have made this journey unforgettable. I am incredibly fortunate to have had you as my supervisor.

Dr. Dana Hayward, our collaboration during the pandemic was a silver lining in an otherwise isolated time. Your brilliance and dedication to research have had a profound impact on me. To my committee, Dr. Kyle Mathewson, and previously, Dr. Kelvin Jones, thank you for your time and insights. Dr. Carrie Demmans Epp and Dr. Mohamed Khamis, your examination contributions are greatly appreciated.

To my lab-mates and academic colleagues in the ACELab and beyond, you've made the past eight years so enjoyable. Nathan and Scott, you've been my steadfast companions through it all. Your support, both scientific and personal, has been invaluable, and your friendship (including the countless Discord calls) has been a highlight of this PhD. Brea, Helen, and Alex, it has been an honor working with such brilliant women. Each of you has enriched me as a scientist and as a person, and I'm grateful for our shared experiences. Many other ACELab members, past and present, have also left their mark on my journey. A special thanks to the students who volunteered and worked in the lab supporting our research.

I'm grateful to the University community, particularly the KSR faculty, for their assistance in navigating academic processes. Special thanks to Elisha Krochak, the invaluable KSR Grad Program Coordinator. I also want to thank the KSR faculty, NMHI, and Psychology Department for fostering vibrant environments for research and collaboration. My time in these halls has been a privilege. I'm particularly grateful to Dr. John Lind, whose teaching ignited my enthusiasm for advanced statistics (against all odds).

I extend my gratitude to the generous funding sources that supported my graduate studies, including the SMART Network, AGRI, Mitacs, and the University of Alberta. Furthermore, I am immensely thankful for the opportunities and experiences this PhD journey has afforded me, including summer courses at Lethbridge and McGill, memorable conferences in Europe, San Diego, and elsewhere, and a rewarding Mitacs Accelerate International internship. Special thanks to Holger Finger and Caspar Goeke from Scicovery GmbH for supporting my internship.

In conclusion, I am eternally grateful for the privilege of pursuing my PhD and creating this thesis. Although my name appears as the sole author, it is a collective effort that I could dedicate an entire thesis chapter to. Thank you.

Table of Contents

1 Introduction **1**

1.1 Preamble . . . 1

1.2 Eye-tracking . . . 5

1.2.1 Modern video-based eye-tracking . . . 8

1.2.2 Webcam eye-tracking . . . 15

1.3 Remote & Online Experimentation . . . 20

1.3.1 Background - What it is and how it works . . . 21

1.3.2 Advantages and disadvantages . . . 23

1.4 Practical Implications of Remote Webcam Eye-tracking . . . 29

1.4.1 Challenges that need to be overcome . . . 30

1.4.2 Testing the utility of webcam eye-tracking with practical applications in two domains . . . 31

1.5 This Thesis . . . 35

1.6 References . . . 39

2 Assessing webcam eye-tracking utility in digital object interactions **50**

2.1 Introduction . . . 51

2.2 Related Works . . . 53

2.2.1 Real-World Object Interactions . . . 53

2.2.2 Lab-based Digital Interactions . . . 54

2.2.3 UX-Focused Digital Interactions . . . 55

2.2.4 Webcam Eye-tracking . . . 56

2.3 Methods . . . 57

2.3.1 Participants . . . 57

2.3.2 Materials . . . 58

2.3.3 Digital Task Layout . . . 58

2.3.4 Task . . . 59

2.3.5 Procedure . . . 60

2.3.6 Data Processing . . . 62

2.3.7 Segmentation 64
2.3.8 Data Analysis 66
2.4 Results 68
2.4.1 Online, webcam eye-tracking can be a suitable method for quantifying gaze behaviours 68
2.4.2 Digital object interactions yield unique, context-specific movement dynamics 70
2.4.3 Eye-cursor coordination during Pick-up interactions resembles the real world, while Drop-off coordination flexibly conforms to the digital context 71
2.5 Discussion 74
2.6 Limitations and Future Directions 79
2.7 Conclusion 82
2.8 References 84

3 Assessing webcam eye-tracking utility in binary choice decision-making 88
3.1 Introduction 89
3.2 Results 93
3.2.1 Response Time: Hard decisions take longer than easy decisions 94
3.2.2 Proportion of Trials: Unique task demands drive unique gaze behaviours 94
3.2.3 Gaze Dynamics: Driven early by stereotyped information gathering, affected later by decision difficulty 101
3.3 Discussion 107
3.4 Limitations and Conclusion 115
3.5 Methods 116
3.5.1 Participants 116
3.5.2 Materials 117
3.5.3 Task & Procedure 117
3.5.4 Data Processing 120
3.5.5 Dependent Measures 123
3.5.6 Statistical Procedure 125
3.6 References 127

4 A practical guide to webcam eye-tracking 131
4.1 Introduction 131

4.1.1 Background . . . 132
4.1.2 Organization of this guide . . . 134
4.2 Before Data Collection . . . 135
4.2.1 Experimental design within the constraints of webcam eye-tracking . . . 136
4.2.2 Opportunities to work around or within the constraints of webcam eye-tracking . . . 139
4.2.3 Testing your design with consideration of the constraints of webcam eye-tracking . . . 143
4.2.4 Other considerations before collecting data . . . 145
4.3 During Data Collection . . . 149
4.3.1 Filtering participants in theory . . . 149
4.3.2 Filtering participants in practice . . . 154
4.4 After Data Collection . . . 157
4.4.1 Assessing raw webcam eye-tracking data quality . . . 158
4.4.2 Data Processing . . . 166
4.5 Summary of Recommendations . . . 172
4.6 Conclusion . . . 172
4.7 References . . . 174

5 Discussion **178**
5.1 Study Summaries . . . 179
5.1.1 Chapter 2 - Assessing webcam eye-tracking utility in digital object interactions . . . 179
5.1.2 Chapter 3 - Assessing webcam eye-tracking utility in binary choice decision-making . . . 183
5.1.3 Chapter 4 - A practical guide to webcam eye-tracking . . . 188
5.2 What is utility anyways? . . . 189
5.3 The future of webcam eye-tracking . . . 191
5.3.1 The pessimist's perspective . . . 193
5.3.2 The optimist's perspective . . . 194
5.3.3 The realist's perspective . . . 196
5.4 My final thoughts: a reflection . . . 197
5.5 References . . . 199

Bibliography **204**

Appendix A: Continuous Measures of Decision-Difficulty Captured Remotely: I. Mouse-tracking sensitivity extends to tablets and smartphones (Ouellette Zuk, Bertrand & Chapman, 2023) 221

Appendix B: Supplemental Materials from Chapter 2 Publication (Bertrand & Chapman, 2023) 242

B.1 README 242
B.2 Prolific Project Description 244
B.3 Data Processing Pipeline 246
B.4 Follow-up Pairwise Comparisons 248
B.5 Spatial Assessment of Clustering Process 251
B.6 Labvanced Study Link 253

Appendix C: Supplemental Materials from Chapter 3 Pre-Print (Bertrand, Ouellette Zuk & Chapman, 2023) 255

List of Tables

3.1 Proportion of trials F-Test results 102
3.2 Pairwise comparisons 108

List of Figures

1.1 Eye Traces from Yarbus, 1967 7
1.2 Components of video-based eye-tracking 9
1.3 Eye detection and localization methods 10
1.4 Examples of eye-tracking hardware 12
1.5 Effects of camera setup and hardware 13
1.6 Applications of remote eye-tracking 15
1.7 Eye images with different eye-tracking hardware 19
1.8 Costs and benefits of online research 27

2.1 Digital task layout, real-world inspiration, task sequence and segmentation approach 59
2.2 Raw eye-data sample 65
2.3 Fixation duration results 69
2.4 Dynamics of eye-cursor coordination 72

3.1 Response Time and Gaze Behaviour Results 95
3.2 Gaze dynamics of dwell patterns 104
3.3 Experimental Paradigm 121

4.1 Designing a webcam eye-tracking experiment 141
4.2 Participant filtering in practice 153
4.3 Timing Variability between eye-tracking samples 162
4.4 Effective sampling rate across study duration 164
4.5 Assessing spatial error with various measures 167
4.6 Task-specific measures for data validation 170
4.7 Summary of recommendations for online webcam eye-tracking experiments 173

Chapter 1

Introduction

1.1 Preamble

It's 2050. You emerge from your catastrophic weather-proof sleeping pod after it gently wakes you up at the perfect moment in your sleep cycle. Energized, you pet your robot dog, whose qualities are so life-like you're still in disbelief that he's not really "real". You reach for a small case on your nightstand, opening it and then popping in your freshly-charged contact lenses, ready to start your day. As you make your way into the kitchen, you glance at the caffeine dispenser, turning on the machine and confirming your usual morning beverage order with a couple quick blinks. Your gaze moves to your temperature-regulating food cabinets, where the opaque cabinet doors become transparent for you to survey their illuminated contents. As you browse, you're feeling uninspired, but unbeknownst to you, a couple items have piqued your subconscious interest, and moments before you become frustrated with your AI-grocer, your robot pup suggests a mangoberry and cocomelon smoothie. He's a genius! It never ceases to amaze you how he always knows exactly what will hit the spot. Sure, the world is simultaneously burning while drowning, and you're confined to a small bunker somewhere in the Canadian mountains, but wow is it ever cool to have a pair of eye-tracking contact lenses that make life so much easier.

Ok, it's 2023 again. While you're maybe a bit apprehensive (or filled with existential dread) about the future, especially for those that believe in science, you have to at least admit that a future with eye-tracking contact lenses sounds pretty neat. In the meantime, what if everyone wore specialized eye-tracking glasses? Besides making our daily activities more convenient with painless Internet of Things device interactions, or enhancing a gaming experience, a world with ubiquitous eye-tracking would likely be much safer. Operating a vehicle or operating on a brain tumor - these are the kinds of riskier moments that could be de-risked with eye-tracking. If everyone wore eye-tracking glasses, there could be multiple solutions to preventing driving accidents, like disabling a vehicle's ignition if the driver displays intoxicated gaze patterns, or alerting nearby pedestrians at a crosswalk that an approaching vehicle is being driven by a person whose eyes are on their phone screen instead of the road.

Of course, no one really owns or wears this specialized eye-tracking equipment, and even in a dystopian (utopian?) 2050 future, it's probably unlikely we'll wear eye-tracking devices, regardless of their form factor. But, what if I told you that right now, upwards of one billion households worldwide may have the components of an easy-to-use and unobtrusive eye-tracker already (Statista Search Department, 2021)? If you missed the title of this thesis, you might be surprised to find out that this magical and pervasive eye-tracking technology I'm referring to is in fact webcam eye-tracking. There's no doubt that the global pandemic accelerated users' comfortability with webcams, with videoconferencing becoming the safest way to engage in teaching kindergarteners, visiting with aging family members, or attending global academic conferences. For most computer owners and users, if they didn't have a webcam prior to the pandemic, it was one of the first purchases for their new home office. For instance, Logitech, a consumer hardware manufacturer, reported modest year over year increases of around 8% for their PC webcam sales until 2021, when the pandemic sent webcam sales into a tailspin. In twelve months, webcam sales alone increased 240%, from $129 million to $440 million, and the next year, another $400 million's worth of webcams were sold at Logitech (*Logitech 2023 Annual Report, SEC Form 10-K*, 2023). That is a lot of new eye-trackers on the market.

Eye-tracking using a webcam is relatively new (first iterations tested by Bäck, 2006), but eye-tracking as a science isn't. For the last 200 years, scientists have sought to understand bigger-picture concepts like human cognition, perception and attention from eye movements. How we look at things, what we look at, when we look at them - chasing the answers to these questions offers a chance to peek inside the inner workings of the brain. Over the years and even now, eye-tracking primarily occurs in the lab with expensive, highly specialized equipment that can measure where the eyes are looking at every half-millisecond with pinhead-level accuracy. These laboratory eye-tracking systems have enabled ground-breaking, foundational research in many domains, but their use, just like all laboratory systems for human research, comes at a cost.

The traditional approach to human research, where undergraduate students attend a controlled lab space to participate in rigid, paradigmatic experiments, has increasingly been challenged for its validity and applicability. While the concept is complex, and has its own criticisms (that are beyond the scope of this thesis, but see Holleman et al., 2020), the 'ecological validity' of research has become an important factor in modern day human research. Put simply, how much of the human experience that's been measured in the confines of a highly-controlled lab environment actually represents the true human experience of the real world? The idea of ecological validity has exposed a tradeoff that it seems flew under the radar for some time, where experimental control has been prioritized at the expense of ecological validity. Then tension between ecological validity and experimental control is itself a moving target, having gone from the relaxed methodologies of single-subject (often self-testing) experiments of the late 1800's and early 1900's to rigorously-controlled experiments in modern university laboratories. But, trying to measure true and real human behaviours as they unfold in everyday life presents all sorts of challenges, from the logistics of recording those behaviours, to the complexities of drawing meaningful, scientific conclusions from undefined and diverse behaviours. What then is the sweet spot?

The idea of tradeoffs, or finding the sweet spot, is central to this thesis - from the experimental process at large to the specific decisions you make implementing a single experiment or analyzing a single dataset.Throughout these thesis projects, the global pandemic enforced a shift to remote, online research methods that pushed the testing environment to a more ecologically-valid one. This was exciting, especially when I came across a remote-friendly method of eye-tracking. Quickly, however, I recognized one of the major tradeoffs that came with remote webcam eye-tracking - sure I could test participants who were comfortable (and safe from covid-19) in their own environments and on their own devices, but I no longer could control that environment. The devices people used were their own, encouraging more authentic device-use, but those devices varied in screen size, and were sometimes missing the resources to power the webcam eye-tracking program. And unlike the lab, where I could engineer an environment free from distractions, I had no idea if participants were highly focused, or were sharing their attention with a TV program in the background. The problem with all of these releases of experimental control (as a result of remote experimentation methods in general) is that they can add up to a collection of datasets that are too noisy or full of extraneous factors for there to be any meaningful conclusions to be drawn from them.

This experimental control challenge emerges with any shift to remote data collection, but when using a method like webcam eye-tracking, there's an additional layer of noise in the eye-tracking data as a result of using remote-friendly methods. Specifically, there emerges a tradeoff in data quality between the non-remote but ultra-precise laboratory equipment and the remote-enabled but consumer-grade webcam. But, is there a "sweet spot" in all of this, where remote webcam eye-tracking affords enough data quality that meaningful, ecologically-valid conclusions can be made?

The primary motivation for this thesis stemmed from the search for this perfect balance. Could we explore the practical value of remote webcam eye-tracking by applying it in human research contexts and assessing whether it can produce data that can contribute to meaningful conclusions about human behaviour? The rapid development of remote webcam eye-tracking has meant its utility and application have been relatively underexplored. Some works have approached the assessment of webcam eye-tracking's utility by performing technical validations on its precision and accuracy (e.g. Burton et al., 2014; Wisiecka et al., 2022), but my approach anchors to the method's usefulness in capturing real human behaviours - assessing its ability to achieve the optimal balance.

The remainder of the introduction presents the reader with some of the core background knowledge required to fully understand the significance of the contributions of the three thesis studies. In order to appreciate the pervasive and nuanced challenges of webcam eye-tracking, it is first useful to provide an overview of eye-tracking in general. I explain how it works in the typical lab context, and then discuss how the eye-tracking process becomes more challenging when lab-grade eye-trackers are exchanged for consumer-grade webcams. I then present an overview of online and remote experimentation methods and discuss the pros and cons of trading the lab environment for the accessible remote world. Finally, I join the two topics together, explaining the practical challenges in collecting remote webcam eye-tracking, and explaining the approach I took in exploring whether these challenges could be overcome to still afford utility in the method. I conclude the introduction with high-level summaries of the studies in this thesis.

1.2 Eye-tracking

The study of eye movements dates back at least two centuries, with early researchers like Bell (1823) observing eye movement patterns and their effects on visual orientation. Early eye-tracking methods involved mechanical devices that directly measured movements from the eyeball, such as a plaster contact lens (used with a local anesthetic of cocaine to cause only "a little temporary inconvenience"; Delabarre, 1898) or a mirror-linked rubber suction cup (Yarbus, 1967). Other methods relied on the observation of corneal reflections onto a film using prisms (Buswell, 1935) or electrodes placed around the eye (Young & Sheena, 1975).

Beyond these early methodological innovations, it was the seminal work of Yarbus (1967) that sparked an ongoing pursuit of understanding the brain through eye measurements, quantifying the attentional value of recording gaze patterns while participants observed images. In specific, Yarbus (1967) was the first to show that differing the goal of image viewing produced remarkably different fixation patterns. For example, when presented with an image of a scene of people, fixations would cluster on the faces when asked to guess their ages, or, when asked to assess the people's wealth, participants would fixate on their clothes and the contents of the environment (see Figure 1.1). With these objective measures, Yarbus (1967) identified patterns of fixation that reflected the informational value of different image elements, linking the brain, or as he described it, the "thought process", to the order and duration of fixations. This "eye-mind link" (Rayner, 2009; Rayner & Reingold, 2015) has since been explored beyond visual attention, offering an informative, moment-by-moment window into perception, memory, language and decision-making (Carter & Luke, 2020).

In the following subsections, I describe modern day approaches to eye-tracking (all of which do not require the use of cocaine!). Subsection 1.2.1 (Modern video-based eye-tracking) presents an overview of the primary form eye-tracking takes today, where videos of the eyes are used. The core components of video-based eye-tracking are eye localization and gaze estimation, and I describe some of the different techniques to solve these image processing challenges. In subsection 1.2.2 (Webcam eye-tracking), I discuss the emergence of webcam eye-tracking as a specific form of video-based eye-tracking that makes it possible to use video-based eye-tracking beyond the laboratory. In the context of this thesis, having an understanding of the complexities of the eye-tracking process in a more broad sense allows for a greater appreciation

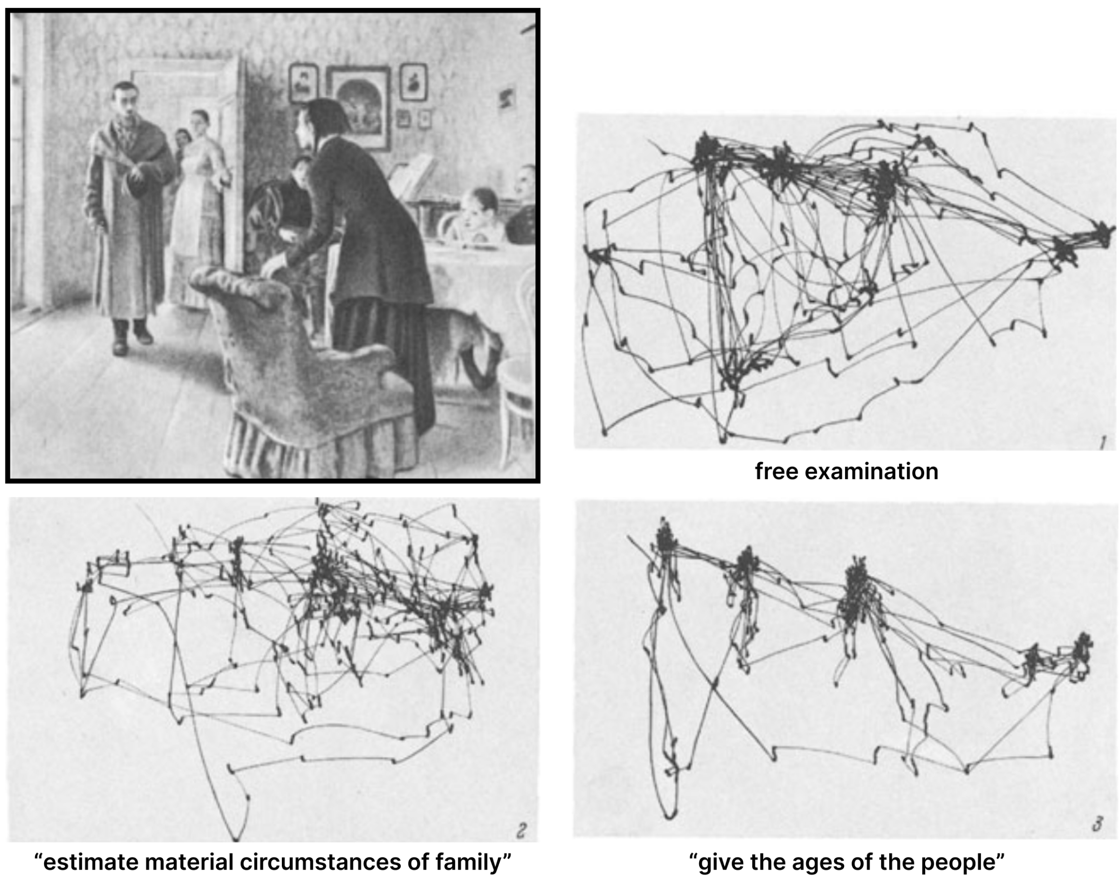


Figure 1.1: Eye movement patterns when viewing an image (photograph, top left panel) with different goals (as stated) below each eye movement trace. Adapted from Yarbus (1967).

of the significant tradeoff that is required when a remote-friendly, consumer-grade webcam is used. I present a broader picture of the various techniques to solving the eye localization and gaze estimation problem as a means of highlighting the flexibility that exists within the method, suggesting that there's certainly opportunities to strike the perfect balance and find the sweet spot.

1.2.1 Modern video-based eye-tracking

Modern day eye-tracking no longer requires intrusive devices, instead relying on more comfortable video-based techniques. This involves capturing images of the eyes frame by frame using a camera and light source, or multiple cameras and/or lights. Hansen and Ji (2010) and later Cristina and Camilleri (2018) describe video-based eye-tracking as a process involving two interconnected components or domains, as depicted in Figure 1.2. The first component is eye localization, which involves detecting the eye within a camera image and consistently monitoring its position across multiple frames. The second is gaze estimation, which requires detecting and tracking the eye while also incorporating the head's position to accurately determine the location of the gaze.

The first component of video-based eye-tracking, eye detection, has been tackled with a variety of approaches for identifying the eye from video images. Model or shape-based methods are often used in conjunction with infrared cameras, which can produce very clear, high-contrast images of the eye (see Figure 1.3). These methods detect the eye by the iris and pupil's stereotypic circular or elliptical shapes, as well as the shape of the eyelids, doing so by fitting a deformable shape template to the infrared camera images of the eyes (Hansen & Ji, 2010). Because the deformable shape model is naturally flexible (i.e. deformable), model or shape-based methods of eye detection are capable of handling frame to frame changes in the eye's shape, scale, and rotation (Hansen & Ji, 2010). In contrast, appearance-based methods are primarily applied to eye images taken under indoor or ambient lighting conditions, where the

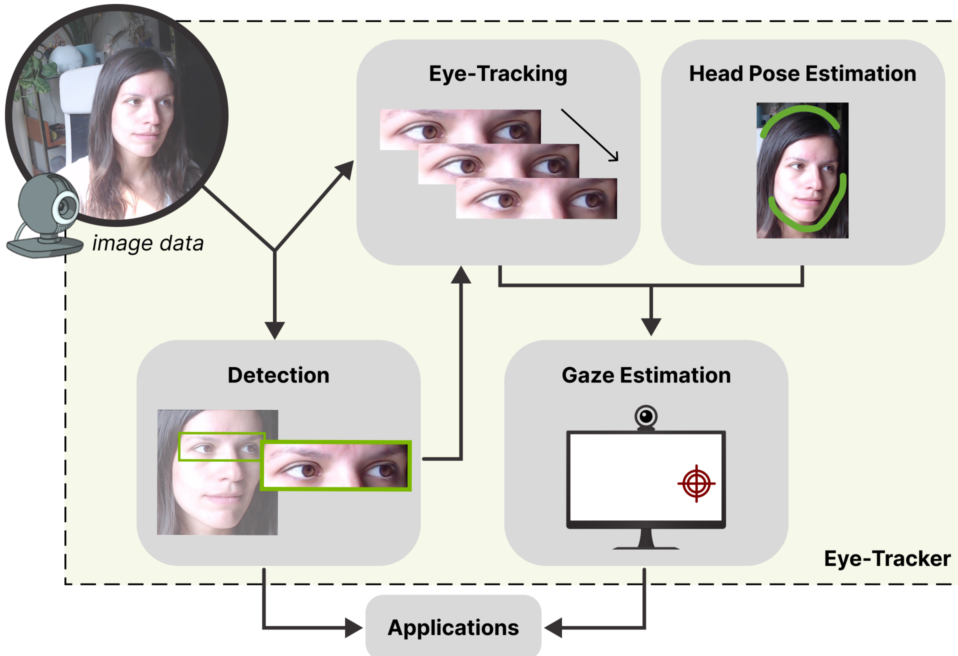


Figure 1.2: The components of video-based eye detection and gaze tracking, adapted from Hansen and Ji (2010).

way the visible spectrum light reflects off the eye (i.e. the photometric properties of the eye) facilitates the eye's detection (see Figure 1.3). These methods are independent of geometry, instead applying image processing techniques that use templates or distributions of pixel intensities to detect the eye within the image. Generating these templates is typically accomplished with machine learning, but substantial training data is required, making appearance-based methods more adversely affected by frame-to-frame head and eye movements where the eye's changed appearance hasn't been trained. Another approach to eye detection is feature-based methods, which focus on identifying specific local features of the eye (as opposed to the entire eye model like model/shape-based methods; Hansen and Ji, 2010; see Figure 1.3). These methods employ image processing techniques that attempt to identify a feature like the limbus or pupil by finding the feature's bounds with intensity differences or filter responses. Because this approach compares relative intensities within each image, feature-based methods are more forgiving of lighting variability or head movements. All of these methods, usually separately but sometimes in hybrid combination approaches, are

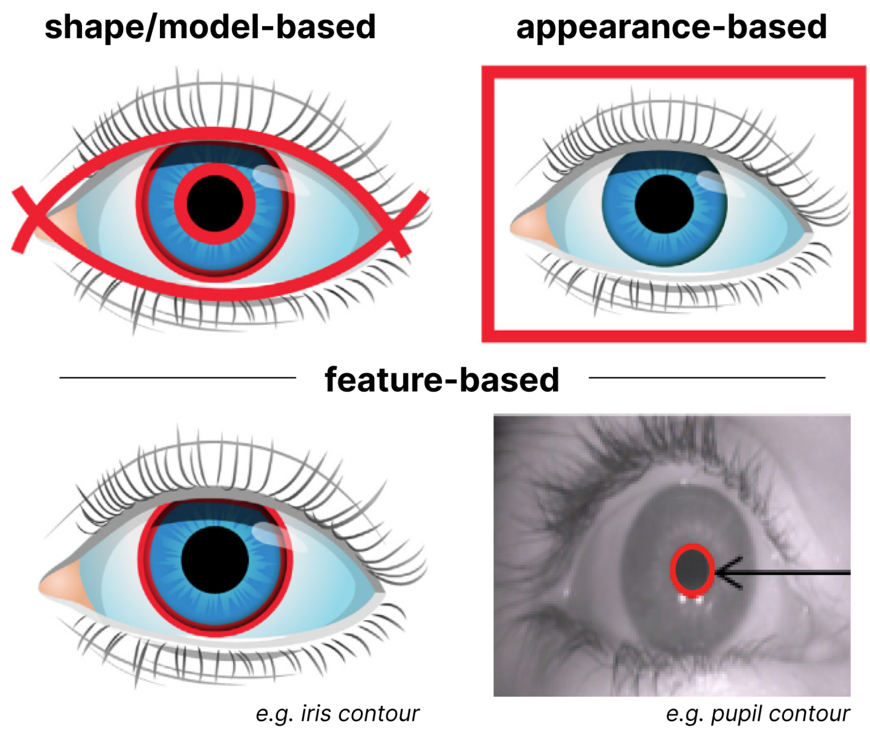


Figure 1.3: Eye detection and localization methods. Adapted from Cristina and Camilleri (2018) and Shehu et al. (2021b).

used as the first step in various contemporary, video-based eye-tracking solutions.

However, localizing the eye in a frame of video data is only the first step in determining where that eye is looking. The output of the second component, gaze estimation, is what this thesis is most concerned with. Gaze estimation requires the derivation of a set of parameters to best estimate the direction of the gaze from the image data (Hansen & Ji, 2010). Various calibration procedures are essential for determining or acquiring these parameters and, in turn, achieving accurate gaze estimations. Some calibrations are performed only once, like when a new camera is used, where setup and camera calibrations are performed to ensure the imaging hardware's technical specifications are considered in the gaze estimation predictions (Cristina & Camilleri, 2018; Hansen & Ji, 2010). Other calibrations are performed for each new testing session. Gaze-mapping calibrations define the parameters required for translating eye data into the coordinate space of the head and camera, and point-of-regard calibrations further extend the gaze information onto a computer monitor or planar surface (Cristina & Camilleri, 2018). In the same fashion as feature-based eye detection models, feature-based gaze estimation models also use the extracted local features of the eye image, this time in service of mapping the relationship between those features and the gaze's position. This mapping is founded on the initial gaze-mapping calibration procedures, but is further built out with an interpolation approach (or other more advanced approaches). Beyond feature-based gaze estimation models, appearance-based models of gaze estimation have also started to emerge as a solution to lower resolution input images (Hansen & Ji, 2010).

Clearly the quality of the image data is crucially important to both components of eye-tracking, so it follows that the camera used to capture those eye images is a vital component of the equation. Achieving optimal image data quality has been primarily accomplished through the use of specialized equipment, such as infrared light imaging and careful camera orientations. Infrared light cameras provide high-contrast images of the eyes by using controllable yet invisible infrared lighting, overcoming the variability that comes with visible spectrum lighting (Hansen & Ji, 2010). For tasks completed on the computer screen, infrared cameras placed near or embedded within the screen have been commonly used in laboratory settings. Traditionally, eye trackers like the commonly-used EyeLink 1000 (SR Research, Ottawa, Canada; see Figure 1.4) are used in conjunction with some form of physical restraint, like a bite bar or a chin rest, in order to minimize the impacts of head movements on the eye tracking process (seen in Figure 1.4). However, more recently, there's been a move towards cheaper, more lightweight options like the screen-mounted Tobii Pro Fusion Eye-tracker (Tobii Research AB, Sweden) or the Gazepoint GP3 eye-tracker (Gazepoint, Vancouver, Canada). Another hardware option is placing infrared cameras directly on the head, eliminating the challenge of head movements, but introducing a different challenge of situating the head's position (i.e. gaze origin) in 3D space (Stone et al., 2022). While older models may have been cumbersome and inhibited movements, today's head-mounted eye-trackers like Tobii's Pro Glasses 3 (Tobii Research

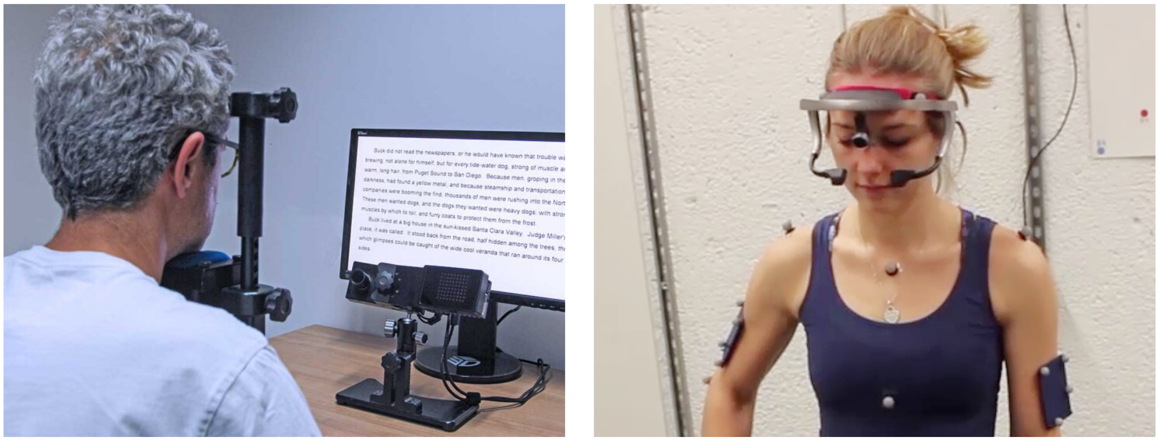


Figure 1.4: Examples of eye-tracking hardware. The left panel shows a desktop-mounted EyeLink 1000 system where a chinrest is used to keep the head stationary. The right panel shows a head-mounted Pupil Labs eye-tracker, used in experiments by Lavoie et al. (2018). Left image from SR Research (www.sr-research.com/eyelink-1000-plus/) and right image adapted from Lavoie et al. (2018).

AB, Sweden) or Pupil Labs' Pupil Invisible (Pupil Labs GmbH, Berlin, Germany) are no more intrusive than a pair of glasses, making them suitable for tasks that involve physical activity or aren't confined to a computer screen, such as Lavoie et al. (2018) pictured in Figure 1.4.

With all the different ways to capture images of the eyes, the temporal and spatial resolution of the recorded video images is an important factor in accurate eye detection and gaze localization. In this image data context, spatial resolution pertains to the number of pixels contained within the captured image, whereas temporal resolution refers to the frequency at which images are captured over a period of time. In the case of lab-grade tools like the EyeLink, Tobii and GazePoint, these specialized pieces of hardware offer the temporal resolution of hundreds to thousands of samples per second and use cameras with the spatial resolution to capture changes of 0.01° (root mean square; *EyeLink 1000 Plus Technical Specifications*, 2017). Importantly, the camera's placement, especially in terms of its proximity to the eyes, affects the spatial resolution of the eye image. This was explored by Kar and Corcoran (2016), who investigated the effects of camera resolution and user distances from the camera on gaze estimation

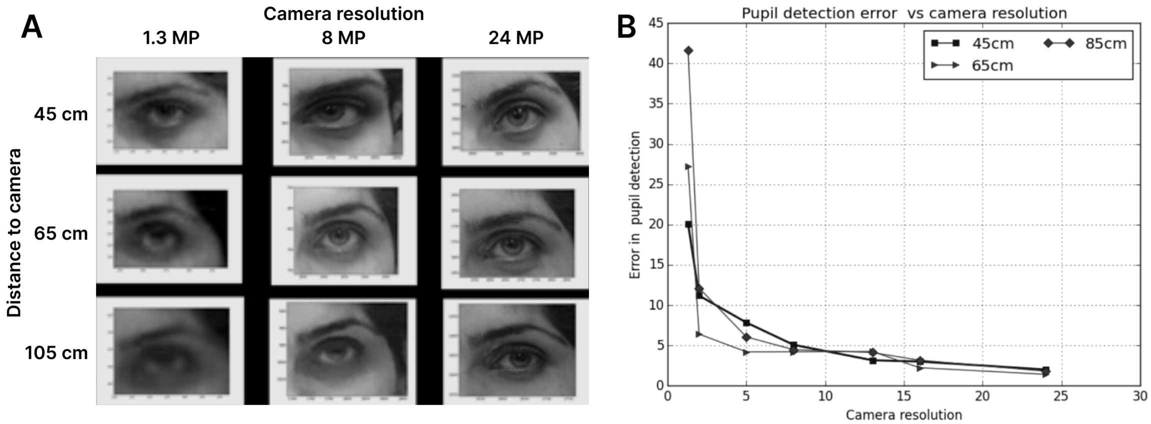


Figure 1.5: Effects of different camera setups and hardware. A) Images of the eye using different camera resolutions at different distances from the user. B) Pupil detection errors as a function of camera resolution and distance from the user. Pupil detection errors are measured as the positional offset (in pixels) between the algorithmically-determined pupil center and the true pupil center. Adapted from Kar and Corcoran (2016).

errors in ambient light settings (see Figure 1.5). They found that more distance between the user and camera produced greater errors in lower resolution cameras, and that pupil detection errors reduced substantially with cameras of a higher resolution (8MP or higher; Kar and Corcoran, 2016). Again, this work was using ambient light cameras, but it's worth noting that the type of illumination (infrared vs ambient light) plays a significant role in the spatial resolution the eye-tracking system requires to make accurate predictions. With the high contrast imaging of infrared cameras, feature extraction of the pupil, for example, is likely to require much less spatial resolution than the less-contrasted ambient light equivalent (like the image in Figure 1.5, for example). Returning to the temporal domain, while lab-grade eye-trackers can collect at a 2000 Hz sampling rate, it's been determined that at least 100 samples per second (i.e. 100 Hz sampling rate) are required to capture high-frequency eye movements or saccades without motion blur artifacts or missing saccades (Abbott & Faisal, 2012; Kar & Corcoran, 2016).

While enhancing the spatial and temporal resolution of eye image data would generally improve the signal to noise ratio and in turn increase eye-tracking accuracy, there are more nuanced parts to the story. These nuances harken back to the constant need to balance tradeoffs; in this case it's finding the sweet spot of image data quality. For example, a trade-off emerges between a larger field of view to track head movements and a narrower field of view to capture high-resolution eye images (Hansen & Ji, 2010). Further, depending on the use case, the additional computational resources required to process higher spatial and temporal resolution image data may inhibit real-time processing or affect additional experimental computations and processes. While these factors are in contention with one another, drawing from abstractions of the fourth chapter in this thesis, they can be resolved with methodically testing the tradeoff in order to eventually arrive at a balanced solution.

Interestingly, both components of video-based eye-tracking have utility in a range of applications. Eye detection, for example, is a crucial component in computer vision applications such as facial recognition, facial feature tracking, facial expression analysis, and iris detection and recognition (Hansen & Ji, 2010; Shehu et al., 2021b). On the other hand, gaze tracking serves as more than just an empirical tool for studying real-time cognitive processing. It is used for both diagnostic applications, providing an objective and quantitative measurement of a viewer's point of regard, and interactive applications, where gaze serves as a control input for gaze-contingent tools and gaze-based user interfaces (Hansen & Ji, 2010). Initially applied to enhance accessibility in human-computer interactions with the advent of computers (Levine, 1981), video-based eye-tracking has gained versatility over time. This is due to advancements in hardware and software quality, enabling remote and real-time data capture and expanding its applicability across various domains. Figure 1.6 from Shehu et al. (2021b) highlights emerging eye-tracking applications across fields and deployment platforms including device interactions like IoT smart home control (Klaib et al., 2019), assessing human behaviors like confusion (Salminen et al., 2018), supporting medical image interpretation (Brunyé et al., 2019), and advancing biometric procedures like gaze-touch authentication (Abdrabou et al., 2019; Khamis et al., 2016,

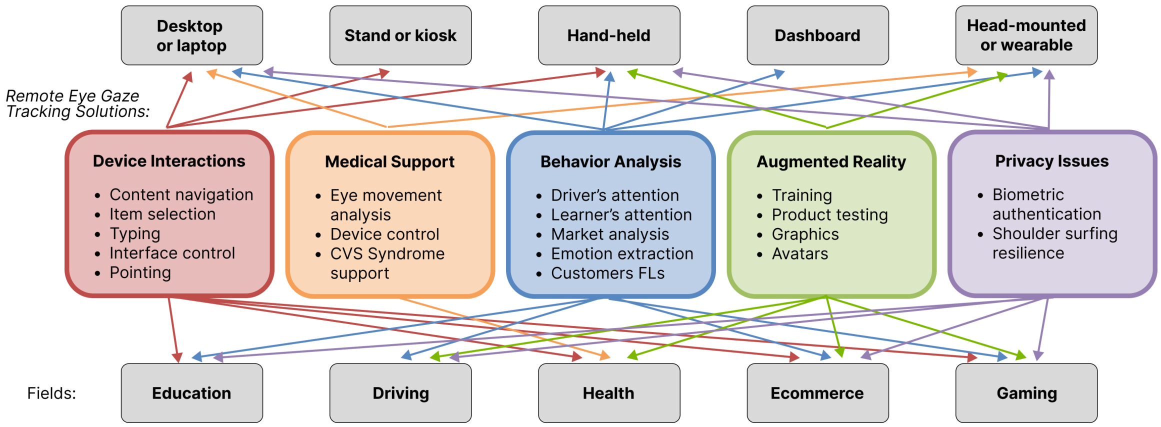


Figure 1.6: Classification of remote eye gaze tracking solutions across fields and deployment platforms. Adapted from Shehu et al. (2021b).

2017).

Now, with an appreciation for the way video-based eye-tracking works, how can we imagine this technology existing beyond the confines of the laboratory? In this thesis, we required a method that afforded remote data capture of the gaze's location. Luckily, in the last decade or so, a remote-friendly form of video-based eye-tracking has emerged, using a work-from-home office staple: the webcam.

1.2.2 Webcam eye-tracking

As video-based eye-tracking evolved, there became an increasing interest in alternative methods to specialized laboratory-grade equipment for capturing eye images. This stemmed from an improvement in both consumer camera technology, making the eye image a higher quality, and the algorithms for eye localization and gaze estimation, making the processes more robust and amenable to lower quality eye images. In particular, the Human-Computer Interaction (HCI) domain was at the forefront of this - excited by the potentially-transformative use of gaze as an accessible input for computer interactions (Levine, 1981). Webcams, already integrated with personal computers and known for their simplicity and low cost, naturally emerged as a suitable candidate for video-based eye-tracking cameras.

HCI researchers were the primary contributors to the early development of webcam eye-tracking as a tool (Burton et al., 2014; Meng & Zhao, 2017; Papoutsaki et al., 2016, 2017; San Agustin et al., 2010; Sewell & Komogortsev, 2010; Xu et al., 2015; Zheng & Usagawa, 2018). From these development efforts emerged a handful of Javascript-based, open-source approaches to webcam eye-tracking, including Webgazer (Papoutsaki et al., 2016) with direct applications for web-browsing, and TurkerGaze (Xu et al., 2015), made specifically to integrate with the crowdsourcing platform Amazon MTurk.

These initial webcam eye-tracking approaches involved eye detection and gaze estimation using feature-based methods (Heck et al., 2023), and unique calibration procedures. For Webgazer, regression models are trained during user interactions when web-browsing, operating on the assumption that the eye and cursor are at the same position during clicking interactions (Papoutsaki et al., 2016). Various eye detection methods are employed to narrow the full field of view web-camera image to the upper half of the face, then to the smallest rectangular bounds that fit the contour of the eye (Papoutsaki et al., 2016). From there, the pupils are located (in two-dimensional, coordinate space), and whole-eye image features (in 120 dimensions) are extracted. Through the self-calibration process of clicking during web browsing, the pupil locations and eye features are mapped to screen coordinates through complex linear regression algorithms, including a ridge regression model. Webgazer's ridge regression model is further enhanced by considering additional known human visuomotor patterns during interaction including the time course of about 200-500 milliseconds for a stabilized target fixation (from Rayner, 1998), and the knowledge that intentional cursor movements correlate strongly with gaze locations (as opposed to a stationary cursor; Hauger et al., 2011). This translates into the incorporation of additional pupil location and eye feature samples from up to 500 ms of time preceding a click (and within a close-enough spatial bound), as well as the cursor's click and movement information variably weighted, where a click position receives a full unit weight while a cursor movement receives a half unit weight that decays over time when stationary (Papoutsaki et al., 2016). While perhaps less sophisticated than the webcam eye-tracking approaches that have since followed (including potentially more robust appearance-based methods such as OpenGaze; Zhang et al., 2019), for their time, tools like Webgazer (Papoutsaki et al., 2016) and TurkerGaze (Xu et al., 2015) offered an exciting first-pass approach to webcam eye-tracking. Even today, the core eye-tracking technology of these first tools is still relevant, with impressive advancements and extensions of the program still being developed (e.g. RealEye; Wisiecka et al., 2022).

Not surprisingly, with the invention of webcam eye-tracking came the commensurate demand for its validation as a tool prior to its use for more empirical work. Thus followed a number of studies focused on assessing the performance, utility, and feasibility of the novel method, often in comparison to lab-based methods of eye-tracking. The utility of webcam eye-tracking has been recently tested in a variety of domains including behavioural, psychological and cognitive science (Bogdan et al., 2023; Schneegans et al., 2021; Semmelmann & Weigelt, 2018; Yang & Krajbich, 2021), online learning research (Hutt et al., 2023; Madsen et al., 2021; Robal et al., 2018; Zhao et al., 2017), linguistics (Slim & Hartsuiker, 2022; Vos et al., 2022), marketing (Schröter et al., 2021), and for clinical optometry applications (Bruno et al., 2023). Further, feasibility studies have been performed to assess whether webcam eye-tracking shows promise as a tool for research with specific target populations including infants (Bánki et al., 2022), older adults living with Alzheimer's Disease (Greenaway et al., 2021), or classrooms with neurodiverse students (Wong et al., 2023). Across these recent works, a central theme emerges: there's cautious optimism about webcam eye-tracking's utility as an eye-tracking method.

The temperance of optimism about webcam eye-tracking is a function of the undeniable tradeoffs that emerged during the above validation studies. These challenges relate back to finding the sweet spot of eye image data quality, where the image must hold enough information to predict the gaze's location. Thus, the introduction of a consumer-grade, ambient light camera, paired with its unrestrictive nature, amplifies many of the general challenges of eye-tracking (as discussed in subsection 1.2.1). These challenges include handling illumination variability and head movements (Cristina & Camilleri, 2018; Shehu et al., 2021b), while contending with greatly-reduced spatial and temporal resolution (Heck et al., 2023; Shehu et al., 2021b). Unlike common infrared camera methods used in the lab where lighting position and intensity is controlled throughout an experiment, webcam eye-tracking introduces variance in the illumination of the eyes (Shehu et al., 2021b). Further, webcam eye-tracking typically doesn't restrict head movements with any additional physical implements like in the lab. These tradeoffs are difficult to navigate, but as suggested earlier, opportunities for lessening the tradeoff impact do exist.

In subsection 1.2.1, I highlighted different algorithmic approaches to detect the eye and estimate the gaze. In the context of webcam eye-tracking, some approaches are better suited than others to handle the unique challenges that arise. This is plain to see in Figure 1.7, where the change to a webcam eye image makes it significantly more difficult to detect specific features of the eye. That being said, this was the approach used by early webcam eye-tracking systems like Webgazer (Papoutsaki et al., 2016; though it might explain why some found the gaze prediction inaccuracy an insurmountable hurdle). Instead, today, alternative machine-learning-trained algorithmic approaches, like appearance-based methods, are employed to counteract the challenge of a webcam-generated image. Here, lower quality images are tolerated because appearance-based methods extract whole-image eye characteristics rather than a specific feature like the pupil (see Figure 1.7; Heck et al., 2023; Shehu et al., 2021b; Zhang et al., 2019). In this way, all the pixel intensities of an eye image like that in Figure 1.7 are stored and mapped to a specific known gaze location (from a calibration process). This therefore requires a substantial training dataset (i.e. calibration procedure) in order to build out a regressive model that is capable of accurately predicting untrained gaze locations from a matrix of pixel intensities. Of course, it

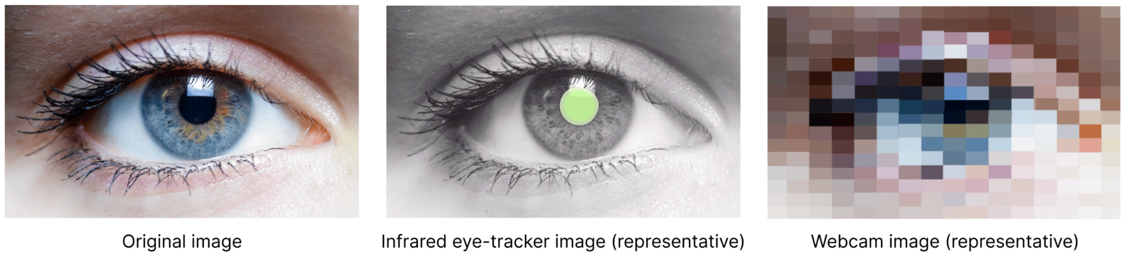


Figure 1.7: A depiction of the eye images captured with different eye-tracking hardware. Adapted from *Webcam-Based Eye Tracking vs. an Eye Tracker [Pros & Cons]*, by O. B. Jensen, 2022 (www.imotions.com/blog/learning/best-practice/webcam-eye-tracking-vs-an-eye-tracker/).

shouldn't then be a surprise that these methods also require navigating a trade-off. For one, this process is computationally demanding, requiring, in our case, computational resources from the remote participant's local device to facilitate the processing of complex algorithms (e.g. convolutional, artificial, or recurrent neural networks; Cristina and Camilleri, 2018; Shehu et al., 2021b). Further, as already stated, a substantial calibration procedure is typically required prior to any eye detection or gaze estimation, though efforts toward efficient and effective calibration are a current focus for some researchers (Gudi et al., 2020; Saxena et al., 2022).

As evidenced, the primary contribution of most webcam eye-tracking research has been methodological in nature, with innovations in the building and validation of the tool rather than in its use. While its novelty effectively necessitates at least some mention of the feasibility or utility of the method across all works, there are a handful of empirically-focused works that have emerged from the adoption of webcam eye-tracking. Recently, webcam eye-tracking has been used to examine visuo-spatial attention in various domains including for object and tool affordances (Federico & Brandimonte, 2019; Federico et al., 2023), face processing (Federico et al., 2021) and body posture preferences (Jacobs et al., 2023). Its use has also been advocated for as a way to remotely assess user experiences (Stone & Chapman, 2023) and the efficacy of clinical treatments (Hsu et al., 2018). In the context of this thesis, Chapters 2 and 3 add to the short list of primarily empirical webcam eye-tracking contributions that have followed the initial methodologically-driven works, however they too provide some assessment of webcam eye-tracking's utility. In contrast, the fourth chapter resembles the earlier methodological works, but takes a more practical approach to the use of the method, while also incorporating considerations for the remote aspect of webcam eye-tracking.

This remote aspect will be discussed next in section 1.3, but is presented first in a more siloed manner to offer a broader picture of the tradeoffs that emerge with remote experimentation methods. Later, in section 1.4, we bring both of these core concepts together, to give context to the unique sweet spot that's required to find utility in remote webcam eye-tracking.

1.3 Remote & Online Experimentation

This thesis centers on the use of a remote tool for capturing complex human behaviours. The shift to collecting data from non-lab contexts affords an opportunity to understand behaviours in the environments that they usually emerge in, rather than the constructed laboratory environment. Methods for remotely capturing simple human behaviors, like survey responses or button presses, have been available for some time. However, the widespread availability of the internet, combined with the global health pandemic, has significantly accelerated online behavioral research across many domains (Bánki et al., 2022; Gagné & Franzen, 2023; Johnson et al., 2021; Weydmann et al., 2022; Yang & Krajbich, 2021). This has led to an increase in the availability of other complementary tools that streamline the research process, as well as the emergence of high-tech platforms that offer no-code experiment building solutions.

In the context of this thesis, our initial motivation for adopting remote methods of experimentation stemmed from the challenges posed by the global COVID-19 pandemic. However, over time, I have come to recognize the remarkable possibilities presented by remote data capture, particularly in its capacity to capture genuine and natural human behaviors in real-world environments rather than controlled laboratory settings. This theme has become central to this thesis, prompting the following exploration into the intricacies of remote data collection (subsection 1.3.1), and the challenges and benefits of transitioning to online platforms from controlled laboratories (subsection 1.3.2).

1.3.1 Background - What it is and how it works

Remote online data collection refers to the use of internet-based tools to gather response data from participants, eliminating the need for their physical presence in a laboratory setting. This method primarily relies on asynchronous participation, allowing individuals to engage in experiments remotely using their personal devices and within their own environments. In the following discussion, we specifically focus on remote data collection methods where participants engage without direct experimenter supervision or moderation. However, it's worth acknowledging that there are innovative solutions for remote yet synchronous data collection in highly-controlled experimental paradigms, such as those used in neuropsychological tests (e.g. Cuttler et al., 2021; Wadsworth et al., 2016).

The process of conducting experiments online, as described by Grootswagers (2020) and Sauter et al. (2020), involves three key components: experiment building, experiment hosting and management, and participant recruitment (a note: these are the components of conducting the experiment and don't pertain to the steps required after the experiment is conducted - we elaborate on the complexities of the later experimentation process in Chapter 4). The first component of online experimentation is experiment building, which, unlike in-lab experimentation, requires the coding of programs specifically tailored for web browser execution. This task is further complicated by the unique and novel nature of experimentation and the scientific process, making a rigid template or a one-size-fits-all approach less useful. Fortunately, the recent availability of GUI-based and low-code experiment building tools has empowered non-web developers, such as human behavioural researchers, to easily create and deploy complex online experiments. A variety of options have emerged in the last decade including: Labvanced (Finger et al., 2017), Gorilla (Anwyl-Irvine et al., 2020), Lab.js (Henninger et al., 2022), jsPsych (de Leeuw, 2015; Leeuw et al., 2023), PsychoPy (Peirce et al., 2019), Empirica (Almaatouq et al., 2021), nodeGame (Balietti, 2017), psychTestR (Harrison, 2020), Pushkin (Hartshorne et al., 2019), OpenSesame (Mathôt et al., 2012), Lookit (Scott & Schulz, 2017), PsyToolkit (Stoet, 2017), and IBEX/PennController (Zehr & Schwarz, 2018). These tools vary in their accessibility (proprietary or open source), functionalities, support, and extensibility. In general, these methods allow researchers to design and implement complex, accurate, and precise online experiments without requiring expertise in web-compatible programming languages like JavaScript.

The second component of online data collection is experiment hosting. Here, in contrast to running an in-lab study where the experimental code and data are stored locally on the testing computer, an online experiment's code and resources must be hosted on the internet through a server (Grootswagers, 2020; Sauter et al., 2020). Commercial experiment-building platforms like Gorilla and Labvanced have built-in server infrastructure that facilitates both executing the experiment (i.e. hosting) and storing the collected participant data. For experiments created with freely available tools like OpenSesame or jsPsych, researchers with advanced technical expertise have the option to use their own web servers or cloud services (e.g. AWS or Google) for hosting, or they can opt for a centralized hosting provider like OpenLab (Shevchenko, 2022) or Pavlovia (www.pavlovia.org; Grootswagers, 2020; Sauter et al., 2020). These services offer hosting solutions designed to simplify study administration, with organized and secure storage of participant data as well as tools that integrate with various participant recruitment platforms.

The final component of online experimentation is participant recruitment (Grootswagers, 2020; Sauter et al., 2020). This stage allows the full potential of online experimentation's efficiency to be realized, as entire experimental datasets can be collected within a matter of hours (A. Newman et al., 2021; Sauter et al., 2020). Participants typically access the study through a provided link and can complete the online experiment as long as they have internet access and a compatible browser. Crowdsourcing platforms like Amazon Mechanical Turk (MTurk; www.mturk.com) or Prolific (www.prolific.co) have been commonly used for recruiting participants in online research. These platforms maintain an active pool of interested participants and offer experimenters simplified participant management, handling of participant compensation, and anonymized, real-time communication channels for troubleshooting technical issues (Aguinis et al., 2021; Johnson et al., 2021). The increasing use of participants from crowdsourcing platforms (Uittenhove et al., 2022) has led to various investigations into the data quality, utility, and population sampling features of these platforms (particularly MTurk; Aguinis et al., 2021; M. Buhrmester et al., 2011; Chmielewski and Kucker, 2020; Crump et al., 2013; Johnson et al., 2021; Paolacci, 2010; Peer et al., 2017; N. Stewart et al., 2017; Thomas and Clifford, 2017). I will explore these aspects further in the next subsection, where I discuss the advantages and disadvantages of online research methods.

1.3.2 Advantages and disadvantages

By now, it should be apparent that online experimentation requires navigating a host of its own unique tradeoffs. At the heart of the transition from tightly controlled laboratory settings to remote online experimentation is the unavoidable reduction of experimental control. This tradeoff of experimental control is perverse and affects all aspects of online experimentation (see Figure 1.8 from Gagné and Franzen, 2023), from the quality of the data collected to the level of participant engagement.

This central experimental control tradeoff primarily manifests in the quality of data collected from remote participants. Lower quality or noisier data arises from variations in participants' hardware and software (Anwyl-Irvine et al., 2020; Pronk et al., 2020; Semmelmann et al., 2017; Uittenhove et al., 2022), and their environments (Clifford & Jerit, 2014; Uittenhove et al., 2022). It is accepted that the accuracy and precision of stimulus presentation and response timings will be less reliable and more variable than those achieved with laboratory-grade equipment (Grootswagers, 2020; Sauter et al., 2020). However, the extent of this challenge appears to be limited, with lab-to-online comparison studies demonstrating minimal and acceptable system variability in stimulus presentation and response timing (e.g. < 10 ms inter-trial variability reported by Bridges et al. (2020) in their comprehensive 'timing mega-study') and successful replication of classic response timing effects (Anwyl-Irvine et al., 2020; Brand & Bradley, 2012; Crump et al., 2013; Gagné & Franzen, 2023; Miller et al., 2018; Pronk et al., 2020; Uittenhove et al., 2022). For example, using crowdsourced online participants, Crump et al. (2013) were able to replicate classic response timing patterns in Flanker, Simon and Stroop paradigms, where incompatible or incongruent items all elicited longer response times than compatible or congruent items. Furthermore, this challenge seems to be diminishing as technology improves and browser display methods are refined for even greater precision (e.g. optimizing the performance of JavaScript timing functions; Garaizar and Reips, 2019; Lukács and Gartus, 2022). However, for more complex forms of online data collection, such as webcam eye-tracking, data quality remains a persistent challenge. For instance, Yang and Krajbich (2021) report excluding only one out of 40 datasets collected in the lab, while needing to exclude half of their online-collected datasets due to poor-quality eye gaze data (though these exclusions occur during initial hardware and calibration checks).

A second tradeoff that emerges with remote data collection is the limited oversight in experimental procedures, making it difficult to ensure participants' comprehensive understanding of the experiment (Chandler et al., 2014; Paolacci, 2010; Thomas & Clifford, 2017) and their maintenance of attention and engagement (Buchanan & Scofield, 2018; Cheung et al., 2017; Clifford & Jerit, 2014; Johnson et al., 2021).

While these challenges can be minimized through well-designed studies that adhere to a growing body of online experimentation recommendations (e.g. the fourth chapter of this thesis, as well as Aguinis et al., 2021; Gagné and Franzen, 2023; A. Newman et al., 2021; Sauter et al., 2020; Thomas and Clifford, 2017), online participants still self-report more environmental distractions than in-lab participants (although it has no effect on task performance, Clifford and Jerit, 2014), and they may multitask during the study (Necka et al., 2016). Moreover, online participants display varied motivations for participating in online research, which can impact study selection, attention and dropouts (Jun et al., 2017). Recent work has also identified between-platform differences in attention, comprehension and dishonesty (Peer et al., 2022), adding further complexity to the situation. Contributions like the fourth chapter of this thesis help experimenters find the sweet spot in these kinds of challenging tradeoffs that come with online experimentation.

A third challenge of remotely collected data stems from the characteristics of the participant samples, which are typically obtained through crowdsourcing or recruitment platforms. While laboratory research is often limited by convenience sampling and may only be generalizable to WEIRD (Westernized, Educated, Industrialized, Rich, and Democratic) participants (Henrich et al., 2010), online sampling is generally more diverse than typical American college samples (M. Buhrmester et al., 2011), and it can help address the WEIRD sampling issue to some extent (Buchanan & Scofield, 2018; Sauter et al., 2020; Uittenhove et al., 2022). However, online sampling still tends to skew towards particular demographic groups (M. Buhrmester et al., 2011; Casler et al., 2013; Johnson et al., 2021; Paolacci, 2010; Peer et al., 2017) and may become less diverse due to events like the covid-19 pandemic, which limited internet access for low-income and minority communities (Lourenco & Tasimi, 2020). Moreover, sampling bias may arise from participants self-selecting into the sample (Aguinis et al., 2021; Cheung et al., 2017; Clifford & Jerit, 2014), and their non-naivety as they become familiar with research tools and experimental paradigms

(Chandler et al., 2014; DeVoe & House, 2016; N. Stewart et al., 2017). Attrition rates of crowdsourced participation in online experiments are also higher than in-lab studies (M. D. Buhrmester et al., 2018; Gagné & Franzen, 2023; Peer et al., 2022; Yang & Krajbich, 2021; Zhou & Fishbach, 2016), with a recent review of MTurk studies with attrition rates ranging from 31.9 to 51% (Aguinis et al., 2021), which can lead to confounded and false conclusions (Zhou & Fishbach, 2016).

As the use of crowdsourcing platforms like MTurk has become popular in academic research, the associated challenges have been extensively explored. This increased popularity has also brought attention to some accessory or tangential challenges that researchers have had to address, such as the risk of collecting fraudulent data from bots (Dupuis, 2019; A. Moss et al., 2021; A. Newman et al., 2021), or, more recently, the pervasive use of large language models like ChatGPT by crowdsourced participants in text summarization tasks (which presents additional concerns when those types of tasks have previously been considered the gold standard for testing the validity of those very models; Veselovsky et al., 2023). Ethical concerns have also been raised about the potentially exploitative and unfair rates of payment (i.e. below minimum wage) typically provided to crowdsourced participants (Lovett et al., 2018; A. J. Moss et al., 2023). Fortunately, with increased awareness of these problems, and practical guides like the fourth chapter, researchers can account for and mitigate these issues by implementing sound and just online experimental procedures and designs, including fair payment (A. J. Moss et al., 2023; A. Newman et al., 2021; Sauter et al., 2020).

Despite the challenges, remote data collection offers several advantages. Online, crowdsourced participants are more diverse than in-lab research populations, including individuals that are traditionally underrepresented in academic research (Aguinis et al., 2021; Bader et al., 2021; Berinsky et al., 2012; M. Buhrmester et al., 2011). Remote data collection also provides accessibility to individuals who may face barriers in accessing traditional lab spaces. Additionally, online methods allow researchers to

COSTS

- Limited control of testing environment, higher risk of distractions
- Limited possibility for intervention during testing in mass online testing, physical absence of researchers
- Noisy data
- Increased lack of motivation and attention due to extended general computer time usage
- Higher dropout rates
- Compensation of cognitive or perceptual differences by expensive hardware
- Varying computer processing capabilities, timing inaccuracies for brief stimulus display
- Greater potential for cheating (providing invalid data) and participant fraud (pretending to be someone one is not)
- Potentially greater temptation to pay participants non-adequately, no in-person interaction
- If a bug is present and data was not collected in batches, much data needs to be discarded
- Reliable access to online studies required (internet access and proper equipment)
- Sampling of non-naïve participants (particularly via platforms)

BENEFITS

- No physical presence needed
- Time reduction for experimenter(s)' testing time commitment – without supervision requirements
- Rapid data collection (in parallel)
- Easily collect more data for larger sample sizes
- Increased possibilities for recruitment, generally more accessible for most people (depending on socioeconomic background)
- Collect more representative and heterogenous samples (depending on sampling method)
- Increased autonomy for participants about the time and location of participation
- Reduced social pressure and feelings of obligation to finish a study
- Access for more trainees to run their own study earlier in their career
- No lab equipment and space needed; all-in-one solutions provided by platforms
- Potential for increased data anonymity (e.g., for special populations)
- Reduced equipment and research costs

Figure 1.8: A list of the costs and benefits of online behavioural research. Adapted from Gagné and Franzen (2023).

access specific or niche participant pools, such as hard-to-reach populations like infants (Bánki et al., 2022), non-medical anabolic steroid users (J. Cohen et al., 2007), or cohorts from different global regions for cross-cultural research (Van Doorn et al., 2017).

One significant benefit is the parallel nature of data collection, allowing for efficient and cost-effective scaling of human data collection (Semmelmann et al., 2017). Unlike physical lab spaces that collect participants individually, online methods enable simultaneous collection of multiple data sets with minimal experimenter oversight (M. D. Buhrmester et al., 2018; Mason & Suri, 2012; Peer et al., 2017; Yang & Krajbich, 2021). This benefit is echoed by many researchers when discussing the strengths of online research, with many noting the data collection period being a matter of hours rather than weeks (e.g. A. Newman et al., 2021; Peer et al., 2017; Yang and Krajbich, 2021). This scalability also means that sample sizes can be much larger with very little additional effort, helping to offset the tradeoff of noisier data by efficiently improving the signal to noise ratio (Aguinis et al., 2021; Gagné & Franzen, 2023; Paolacci, 2010). In addition, these benefits are not at any greater financial cost to the experimenter, making them resource efficient (M. D. Buhrmester et al., 2018; Mason & Suri, 2012; Paolacci, 2010; Peer et al., 2017).

With advancements in technology, online research opportunities have grown, enabling flexibility in more complex experimental design choices (Aguinis et al., 2021; Paolacci, 2010; N. Stewart et al., 2017). Technological improvements, such as precisely timed stimulus delivery and the recruitment of a participant's computer mouse, camera or microphone have enhanced the possibilities of online research (Johnson et al., 2021). Pushing the limits of online experimentation is the recording of rich, continuous forms of data, like the ongoing position of the cursor, or, most relevant to this thesis, the location of the gaze at every moment. In collaboration with my colleagues, we have taken advantage of this opportunity, twice testing the limits of these methods. In a complex, two-player card game, where each player joined remotely, Ma, myself, Chapman and Hayward (2023) successfully recorded the cursors of both participants, and were even able to broadcast their opponent's cursor position to each player. This afforded us an opportunity to explore much more than the simple measures of card selection, revealing changes in cursor movements when the pair of players were playing games cooperatively or competitively with each other. I also collaborated with Ouellette Zuk and Chapman (2023) to produce the initial study of decision-difficulty that my third chapter builds upon. Again, the capabilities of remote data collection were tested by performing a multi-device study, collecting the cursor position from computer users, and the touchscreen finger trajectories from both tablet and smartphone users. Remarkably, rich, high-resolution cursor and touchscreen interaction position data could be recorded from hundreds of remote participants, revealing impressive decision difficulty effects in the trajectory data (see Appendix A for the complete paper).

It's clear that we are at an exciting moment in time, where remote experimentation technology is capable of recording dense, continuous human data. With the parallel advancements in webcam eye-tracking technology it's now possible to remotely deploy gaze-tracking experiments on the internet. This shift to online experimentation is exciting, especially because, despite the change in experimental context, meta-analyses and empirical studies repeatedly find little difference in the validity of online vs in-lab data across various properties and criteria (e.g. A. Newman et al., 2021; Thomas and Clifford, 2017; Uittenhove et al., 2022; Walter et al., 2019; Weydmann et al., 2022). This context shift is also exciting for its use in facilitating the observation of participants' behaviors in their own environments, promoting ecologically-valid research and enabling researchers to explore behaviors that align more closely with natural and realistic everyday activities. This, combined with the emerging opportunities for capturing more complex and continuous forms of behaviour, is what I discuss next in section 1.4.

1.4 Practical Implications of Remote Webcam Eye-tracking

We have now arrived at the intersection of eye-tracking methods and remote online experimentation methods that are the crux of this thesis. It's clear that both of these methods offer numerous opportunities, and that when combined, they have the potential to offer meaningful insights about real-world dynamic human behaviors. However, I have also belabored the significant tradeoffs that each method demands. In this section, I present remote webcam eye-tracking in terms of its own distinct tradeoffs when it is practically applied as a research method. The first subsection acts as a high-level account of the key challenges, and is thus a brief summary of a portion of the fourth chapter. The second subsection serves to explain and setup our approach to assessing the utility of remote webcam eye-tracking. It explains how the second and third chapters are practical applications of webcam eye-tracking, where we understand if it's useful by how well it can capture key behaviours from the eye-hand coordination and decision-making domains.

1.4.1 Challenges that need to be overcome

The implementation of webcam eye-tracking in remote online experiments poses numerous challenges stemming from the various challenges already inherent to the methods of webcam eye-tracking and remote experimentation that exist independently of their combination. The limitations of remote webcam-based methods include compromised spatial and temporal data quality compared to traditional laboratory eye-trackers. Spatial errors can range from 0.88° to 7° (Saxena et al., 2022; Semmelmann & Weigelt, 2018; Shehu et al., 2021b; Skovsgaard et al., 2011), and are often not of a uniform magnitude across different portions of the screen (Semmelmann & Weigelt, 2018; Vos et al., 2022). In previous remote webcam eye-tracking studies, the hardware's internal sampling limits (usually 30 or 60 Hz), paired with the computational demands of the remotely-deployed eye-tracking system, produces a sampling rate in the range of 11.5 Hz to 40 Hz (Bánki et al., 2022; Semmelmann & Weigelt, 2018; Stone & Chapman, 2023; Vos et al., 2022; Yang & Krajbich, 2021).

While even more typical forms of online experiments like surveys can be challenging for participants, deploying webcam eye-tracking over the internet demands considerable effort from participants, requiring hardware, sustained attention, and controlled environmental elements like lighting (Bánki et al., 2022; Gagné & Franzen, 2023; Semmelmann & Weigelt, 2018) (5, 14, 45). This manifests in substantial rates of participant attrition (Yang & Krajbich, 2021). To address these challenges with participants, thorough experimental design, pilot testing, and effective participant communication are essential. In particular, remote webcam eye-tracking demands robust, engaging instructions and the experimenter's virtual availability for responsive troubleshooting.

Finally, for reasons including those above, challenges emerge when it comes time to interpret remotely collected webcam eye-tracking data. For instance, with limited spatial accuracy, recent studies suggest that remote webcam eye-tracking can accurately distinguish between a maximum of 4 to 6 areas of interest on a computer screen (Slim & Hartsuiker, 2022; Yang & Krajbich, 2021). Moreover, slower sampling rates prevent the detection of rapid eye movements, such as saccades and microsaccades, which means webcam eye-tracking is not currently compatible with applications like those that predict emotional and cognitive states from microsaccades (Heck et al., 2023).

Again, the fourth chapter of this thesis provides a more robust and comprehensive view of these challenges, and also elaborates on the solutions that contribute to finding the sweet spot of high-quality remotely captured webcam eye-tracking data. It does so while presenting data-informed lessons learned from the empirical studies I describe in Chapters 2 and 3 of this thesis. Through these two projects, I grew to appreciate and understand first-hand the various challenges faced by a cognitive scientist deploying webcam eye-tracking, and in the next subsection, I explain how these challenging projects were exercises in testing the utility of remote webcam eye-tracking for behavioral research.

1.4.2 Testing the utility of webcam eye-tracking with practical applications in two domains

The purpose of this thesis is to explore the experimental utility of webcam eye-tracking. That is, despite the challenges presented, can webcam eye-tracking be a useful method for understanding human behaviours? In order to understand its utility, rather than taking the approach of a highly technical validation study, or quantifying the accuracy and precision of the tool, I applied the tool to two unique research domains. These domains were selected for their history of producing stable and reliable continuous forms of behavior as measured with well-established tools (some of which we've previously employed in our own lab). Given this, I wanted to explore the way these rich, continuous behaviours might translate (or not) to experimentation with remote webcam eye-tracking. Said a different way, my approach to understanding the utility of this novel method was to apply it in non-novel domains and look for specific, predicted outcomes. The following outlines the selected domains and explains which established, domain-specific, continuous behaviours make them ideal test cases for exploring the utility of remote webcam eye-tracking.

Eye-hand coordination in digital object interactions

The first domain that offers an opportunity for assessing the utility of remote webcam eye-tracking is eye-hand coordination. Research on eye-hand coordination during real-world object interactions has revealed remarkably consistent patterns of behavior. Early works by Land and Hayhoe (2001) and Land et al. (1999) examined tasks such as making tea and preparing a sandwich, where participants displayed consistent eye-hand coordination patterns even in complex and unstructured settings. In these tasks, the gaze exclusively focused on task-relevant objects in the scene and disregarded irrelevant objects. Most prominently, the pattern most commonly observed was one where the eyes led motor actions, fixating on the object to be manipulated for about half a second before the hand's initial movement towards it. The gaze remained fixated on the object until the completion of the manipulation, and then moved on to the next target without returning. This pattern of eye leading motor action extends to other naturalistic, visually-guided interactions like walking (Land, 2006; Patla & Vickers, 2003), typing (Butsch, 1932), and music playing (Furneaux & Land, 1999), with the eye leading the hand for at least 500 ms (and up to about 1 second) before the initiation of the interaction.

More recently, my colleagues have furthered the understanding of these patterns by more precisely quantifying the timings of eye and hand dynamics using modern motion capture and head-mounted eye-tracking technologies (Lavoie et al., 2018). Participants completed two naturalistic movement tasks that resembled typical kitchen activities like picking up and moving a filled cup across a work area or moving a pasta box from a cupboard to the countertop. Their findings aligned with Hayhoe (2000) and Land et al. (1999), showing that eye-hand coordination in both kinds of object interaction involved fixating on the object at least 500 ms before the interaction began. Within 600 ms of the interaction start, the eyes shifted to look ahead to the next area for interaction. This dominant ∼500 ms eyes-leading-hand pattern has proven highly consistent and stable, including in tasks not just involving reaching, and is therefore a good choice of behaviour to look for with webcam eye-tracking. We expect to witness this 500 ms eye-leading-action pattern when explored in yet another type of task, even if the specific task space is a digital one (as necessitated by webcam eye-tracking). However, as this space is untested, beyond our primary motivation of exploring the utility of webcam eye-tracking by applying it to a digital eye-hand coordination task, we also are in a position to make a novel empirical contribution should webcam eye-tracking indeed exhibit utility. Specifically, this work could provide one of the first quantitative insights about the dynamics of eye-cursor coordination during digital object interactions. In sum, if we find this distinct ∼500 ms pattern during a digital object interaction task by using webcam eye-tracking, we will have a strong case study for the utility of webcam eye-tracking while also making a novel empirical contribution to the eye-hand coordination domain.

Binary choice decision-making

Binary decision-making is another domain that offers an opportunity to explore the utility of webcam eye-tracking as a research tool while also contributing to a better understanding of human visual behaviours. In decision-making research, the emphasis has shifted towards understanding the process of real-world decision-making, prioritizing the dynamics of behavior over mere decision outcomes (Cisek & Kalaska, 2010; Dotan et al., 2018, 2019; Gallivan et al., 2018; Wispinski et al., 2020). A prominent method in this context is the reach-decision paradigm, which involves tracking the movements of the finger or hand as it is used to reach towards one of two choice options on a screen. This method has also been robustly proven to extend from 3D reaching (Chapman et al., 2010a, 2010b; Gallivan et al., 2018) to computerized, 2D reach-decisions using cursor-tracking (Freeman, 2018; Hehman et al., 2015; Stillman et al., 2018, 2020). Across contexts, when coupled with a binary choice, the attraction towards each choice option can be continuously measured and tracked (Dotan et al., 2018, 2019; Stillman et al., 2018). The reach or cursor trajectories formed during decision-making will exhibit a continuum of direct to indirect movements that represent the competition between choice options and the difficulty of the decision (Faulkenberry et al., 2016; Freeman, 2018; Hehman et al., 2015; Koop & Johnson, 2013; Maldonado et al., 2019; Stillman et al., 2018, 2020).

The recent digitized and screen-based binary decision-making study I was part of (Ouellette Zuk et al., 2023; see Appendix A) is an excellent example that demonstrates the very strong and robust effects of decision difficulty on movement trajectories. Ouellette Zuk et al. used three different types of binary choices that have each been previously shown (independently) to sensitively reflect decision-difficulty effects through cursor trajectories: objective perceptual judgements, semi-subjective conceptual judgements, and subjective preference judgements. With all three tasks, Ouellette Zuk et al. explored whether effects from hard and easy decisions would replicate on a computer with a cursor, but also, if they were robust enough to extend to choices made on large and small format touch devices (tablets and smartphones). Despite the differences in the way the choice is enacted (cursor movement vs touch interaction), and other features like the substantial variation in screen size, the difficulty of a decision (hard vs easy) was sensitively reflected in multiple features of cursor and finger movement trajectories alike.

Given the robustness of decision-difficulty effects across multiple forms of binary decision in Ouellette Zuk et al.'s task, and the fact that the task was, in part, performed in a digital, webcam-friendly context (the personal computer condition), this specific task offers an opportunity to test the utility of webcam eye-tracking for gaining an even richer picture of the decision-making process. In specific, we are offered an opportunity to take a task with established Hard and Easy conditions, and ask how useful webcam eye-tracking is in generating indices of decision difficulty. While previous work would suggest that eye movements are affected by decision difficulty in a binary choice task (Krajbich & Rangel, 2011; Krajbich et al., 2010), these works use highly-precise lab-grade eye-tracking tools that model difficulty effects in gazing behaviours on a millisecond by millisecond basis. Thus, if we apply Ouellette Zuk et al.'s decision-making task with proven and robust hard and easy decision contexts, will remote webcam eye-tracking have utility in offering some indication of decision difficulty? If so, we will again have the opportunity to not only provide a test case of the utility of webcam eye-tracking, but also offer novel empirical decision-making findings.

1.5 This Thesis

It should be clear now that remotely-collected webcam eye-tracking is an exciting and promising, albeit underexplored, method for understanding human behaviours. The trade-offs that emerge when trying to use webcam eye-tracking abound: tradeoffs between accessibility and experimenter control, between measures with time or space sensitivity, between meaningful/useful data and accessible collection procedures. But at the heart of these trade-offs is the aim of this thesis: to understand whether remote webcam eye-tracking has utility as a method for understanding human behaviours.

As stated, I approach this aim by exploring remote webcam eye-tracking's utility when applied to two unique domains: eye-hand coordination during object interaction (Chapter 2), and binary choice decision-making (Chapter 3). For eye-hand coordination, webcam eye-tracking will be considered useful as a method if the 500 ms eye-leading-effector pattern is observed via webcam eye-tracking. For binary choice decision-making, webcam eye-tracking will be considered useful if difficulty effects emerge in eye-tracking data. To accompany these demonstrations of webcam eye-tracking's possible utility, I also present a practical methodological guide (Chapter 4) to share the knowledge acquired during the challenging processes of applying webcam eye-tracking to those two domains. This guide serves as an alternative way to understand the utility of webcam eye-tracking in that it fills a gap in the current literature about how to practically use webcam eye-tracking. But, more than just offering some very useful and hard-earned tips, Chapter 4 of this thesis is also infused with data demonstrations drawn from the previous two chapters. This means not only can we offer guidance, but we can also provide an evidence-based prescription with respect to data quality, participant drop-out and experiment costs. With this guide, the practical utility of webcam eye-tracking is expanded (and doesn't require an experimenter to spend years of their life becoming an expert and writing a thesis about the matter!).

To briefly summarize, in Chapter 2, I examine webcam eye-tracking as a tool to explore whether eye-hand coordination patterns during object interactions persist in a screen-based digital context. Working from a familiar place of real-world object interactions and the empirically-established way the vision and motor systems combine to facilitate these, I create a screen-based, digital analog to a tested real-world interaction while integrating webcam eye-tracking. I adapt my colleagues' 'Cups Task' (Lavoie et al., 2018) from a real-world, tabletop cup-transferring activity into a two-dimensional, screen-based task, where circles were dragged and dropped to visual targets in sequence with a cursor. Using webcam eye-tracking, we remotely record the gaze position, complemented by a record of the cursor's position, in order to explore the coordination between the two datastreams. Participants completed 50 trials, repeating a learned sequence of 8 drag and drop circle movements on each trial. Here, we explore the utility of webcam eye-tracking in two ways: 1) we explore whether the quality of webcam gaze data, even when supplemented by required data processing and treatments, is sufficient to even begin exploring eye-cursor coordination, and 2) if sufficient, we want to see if webcam eye-tracking has utility in revealing a 500 ms eye-leading-cursor pattern like the coordination patterns of the real-world.

In Chapter 3, I assess the utility of webcam eye-tracking by applying it to the decision-making domain using a replication and extension of a binary choice study. In a similar vein to Chapter 2, this study also builds off an established foundation, this time from previous work with my colleague (Ouellette Zuk et al., 2023, see Appendix A), and before that, from a series of cursor-tracking binary choice tasks. This study repeats the same experimental procedure as the previous work, but with the addition of remotely-captured webcam eye-tracking. Completing the experiment on a computer, participants made decisions by moving their mouse-cursor from a starting position to their selected choice. The decisions were constructed from three established binary choice tasks: a Sentence Verification task (Dale & Duran, 2011), a Numeric-Size Congruity task (Faulkenberry et al., 2016), and a Photo Preference task (Koop & Johnson, 2013). Each task was designed and analyzed to produce hard and easy trials. With very strong decision-difficulty effects emerging in the cursor movements in the earlier study (Ouellette Zuk et al., 2023), here, we wanted to assess the utility of webcam eye-tracking by exploring whether it might reveal additional decision-difficulty effects. Given the limit of cursor-tracking only affording decision information from movement initiation onwards, could webcam eye-tracking prove useful in illuminating whether any decision difficulty effects emerge before the cursor starts to move towards a choice?

In Chapter 4, I describe the utility of webcam eye-tracking in detail, providing a comprehensive guide to practically employing it as a behavioural research method. As previously stated, this paper is borne out of the challenges I faced and the lessons learned while attempting to glean the highest utility from webcam eye-tracking in the first two studies. While both studies mention methodological challenges and solutions, and include supplementary materials that enhanced webcam eye-tracking's utility, they were anchored first and foremost to their respective domains. In the end, the process of designing, building, testing, and analyzing these webcam eye-tracking studies was a process of maximizing the utility of webcam eye-tracking - trying to make the most out of a challenging method. The fourth paper outlines this process, while sparing budding webcam eye-tracking researchers some of the challenges I faced by providing practical solutions and considerations. The paper covers the complete span of the experimental process, defined by information most relevant to the time before, during and after data collection. Importantly I lean on the data generated from my two earlier studies to infuse this third chapter with evidence-based suggestions in an effort to provide the most realistic and practical picture of experimentation with webcam eye-tracking. Thus, Chapter 4 is a demonstration of and contribution to webcam eye-tracking's utility.

This thesis concludes with Chapter 5, a discussion that revisits the contributions made in Chapters 2, 3, and 4 within their methodological, empirical and broader contexts. I analyze Chapters 2 and 3, assessing their ability to demonstrate the utility of webcam eye-tracking and how this utility informed interesting empirical findings. I discuss utility through a new, more broad lens, an unexpected change in perspective that evolved over the course of this thesis. I further offer reflections on the future of eye-tracking, considering it from pessimistic, optimistic and realistic viewpoints. Lastly, I conclude this thesis by reflecting on my challenging yet rewarding experiences with webcam eye-tracking. By the end, I hope to have convinced you, the reader, that remote webcam eye-tracking has immense utility, and that it is possible to achieve the elusive balance between remotely-collected and high-quality webcam eye-tracking data.

1.6 References

Abbott, W. W., & Faisal, A. A. (2012). Ultra-low-cost 3d gaze estimation: An intuitive high information throughput compliment to direct brain–machine interfaces [Publisher: IOP Publishing]. *Journal of Neural Engineering*, *9*(4), 046016.

Abdrabou, Y., Khamis, M., Eisa, R. M., Ismail, S., & Elmougy, A. (2019). Just gaze and wave: Exploring the use of gaze and gestures for shoulder-surfing resilient authentication. *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 1–10.

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations [Publisher: SAGE Publications Inc]. *Journal of Management*, *47*(4), 823–837.

Almaatouq, A., Becker, J., Houghton, J. P., Paton, N., Watts, D. J., & Whiting, M. E. (2021). Empirica: A virtual lab for high-throughput macro-level experiments. *Behavior Research Methods*, *53*(5), 2158–2171.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407.

Bäck, D. (2006). *Neural network gaze tracking using web camera.* Institutionen för medicinsk teknik. Retrieved June 6, 2023, from https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-5579

Bader, F., Baumeister, B., Berger, R., & Keuschnigg, M. (2021). On the transportability of laboratory results [Publisher: SAGE Publications Inc]. *Sociological Methods & Research*, *50*(3), 1452–1481.

Balietti, S. (2017). nodeGame: Real-time, synchronous, online experiments in the browser. *Behavior Research Methods*, *49*(5), 1696–1715.

Bánki, A., de Eccher, M., Falschlehner, L., Hoehl, S., & Markova, G. (2022). Comparing online webcam- and laboratory-based eye-tracking for the assessment of infants' audio-visual synchrony perception. *Frontiers in Psychology*, *12*, 733933.

Bell, C. (1823). XV. on the motions of the eye, in illustration of the uses of the muscles and nerves of the orbit [Publisher: Royal Society]. *Philosophical Transactions of the Royal Society of London*, *113*, 166–186.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's mechanical turk [Publisher: Cambridge University Press]. *Political Analysis*, *20*(3), 351–368.

Bogdan, P. C., Dolcos, S., Buetti, S., Lleras, A., & Dolcos, F. (2023). Investigating the suitability of online eye tracking for psychological research: Evidence from comparisons with in-person data using emotion–attention interaction tasks. *Behavior Research Methods.*

Brand, A., & Bradley, M. T. (2012). Assessing the effects of technical variance on the statistical outcomes of web experiments measuring response times. *Social Science Computer Review*, *30*(3), 350–357.

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing megastudy: Comparing a range of experiment generators, both lab-based and online [Publisher: PeerJ Inc.]. *PeerJ*, *8*, e9414.

Bruno, A., Tliba, M., Kerkouri, M. A., Chetouani, A., Giunta, C. C., & Çöltekin, A. (2023). Detecting colour vision deficiencies via webcam-based eye-tracking: A case study. *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, 1–2.

Brunyé, T. T., Drew, T., Weaver, D. L., & Elmore, J. G. (2019). A review of eye tracking for understanding and improving diagnostic interpretation. *Cognitive Research: Principles and Implications*, *4*(1), 7.

Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods*, *50*(6), 2586–2596.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? [Publisher: SAGE Publications Inc]. *Perspectives on Psychological Science*, *6*(1), 3–5.

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of amazon's mechanical turk, its rapid rise, and its effective use [Publisher: SAGE Publications Inc]. *Perspectives on Psychological Science*, *13*(2), 149–154.

Burton, L., Albert, W., & Flynn, M. (2014). A comparison of the performance of webcam vs. infrared eye tracking technology. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1437–1441.

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology and perception in art* [Pages: 198]. Univ. Chicago Press.

Butsch, R. L. C. (1932). Eye movements and the eye-hand span in typewriting [Place: US Publisher: Warwick & York]. *Journal of Educational Psychology*, *23*(2), 104–121.

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49–62.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? a comparison of participants and data gathered via amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156–2160.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among amazon mechanical turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112–130.

Chapman, C. S., Gallivan, J. P., Wood, D. K., Milne, J. L., Culham, J. C., & Goodale, M. A. (2010a). Reaching for the unknown: Multiple target encoding and real-time decision-making in a rapid reach task. *Cognition*, *116*(2), 168–176.

Chapman, C. S., Gallivan, J. P., Wood, D. K., Milne, J. L., Culham, J. C., & Goodale, M. A. (2010b). Short-term motor plasticity revealed in a visuomotor decision-making task. *Behavioural Brain Research*, *214*(1), 130–134.

Cheung, J. H., Burns, D. K., Sinclair, R. R., & Sliter, M. (2017). Amazon mechanical turk in organizational psychology: An evaluation and practical recommendations. *Journal of Business and Psychology*, *32*(4), 347–361.

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? shifts in data quality and the impact on study results [Publisher: SAGE Publications Inc]. *Social Psychological and Personality Science*, *11*(4), 464–473.

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, *33*(1), 269–298.
Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? an experimental comparison of data quality in laboratory and online studies [Publisher: Cambridge University Press]. *Journal of Experimental Political Science*, *1*(2), 120–131.
Cohen, J., Collins, R., Darkes, J., & Gwartney, D. (2007). A league of their own: Demographics, motivations and patterns of use of 1,955 male adult non-medical anabolic steroid users in the united states [Publisher: Routledge _eprint: https://doi.org/10.1186/1550-2783-4-12]. *Journal of the International Society of Sports Nutrition*, *4*(1), 12.
Cristina, S., & Camilleri, K. P. (2018). Unobtrusive and pervasive video-based eye-gaze tracking. *Image and Vision Computing*, *74*, 21–40.
Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research (S. Gilbert, Ed.). *PLoS ONE*, *8*(3), e57410.
Cuttler, C., LaFrance, E. M., & Stueber, A. (2021). Acute effects of high-potency cannabis flower and cannabis concentrates on everyday life memory and decision making [Number: 1 Publisher: Nature Publishing Group]. *Scientific Reports*, *11*(1), 13784.
Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science*, *35*(5), 983–996.
Delabarre, E. B. (1898). A method of recording eye-movements [Publisher: University of Illinois Press]. *The American Journal of Psychology*, *9*(4), 572–574.
de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12.
DeVoe, S. E., & House, J. (2016). Replications with MTurkers who are naïve versus experienced with academic studies: A comment on connors, khamitov, moroz, campbell, and henderson (2015). *Journal of Experimental Social Psychology*, *67*, 65–67.
Dotan, D., Meyniel, F., & Dehaene, S. (2018). On-line confidence monitoring during decision making. *Cognition*, *171*, 112–121.
Dotan, D., Pinheiro-Chagas, P., Al Roumi, F., & Dehaene, S. (2019). Track it to crack it: Dissecting processing stages with finger tracking. *Trends in Cognitive Sciences*, *23*(12), 1058–1070.
Dupuis, M. (2019). Detecting computer-generated random responding in questionnaire-based data: A comparison of seven indices, 10.
EyeLink 1000 plus technical specifications. (2017). SR Research. Retrieved July 18, 2023, from https://www.sr-research.com/wp-content/uploads/2017/11/eyelink-1000-plus-specifications.pdf
Faulkenberry, T. J., Cruise, A., Lavro, D., & Shaki, S. (2016). Response trajectories capture the continuous dynamics of the size congruity effect. *Acta Psychologica*, *163*, 114–123.
Federico, G., & Brandimonte, M. A. (2019). Tool and object affordances: An ecological eye-tracking study. *Brain and Cognition*, *135*, 103582.

Federico, G., Ferrante, D., Marcatto, F., & Brandimonte, M. A. (2021). How the fear of COVID-19 changed the way we look at human faces [Publisher: PeerJ Inc.]. *PeerJ*, *9*, e11380.

Federico, G., Osiurak, F., Brandimonte, M. A., Salvatore, M., & Cavaliere, C. (2023). The visual encoding of graspable unfamiliar objects. *Psychological Research*, *87*(2), 452–461.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: A unified JavaScript framework for online studies. *International Conference on Computational Social Science*, *1*(1), 1–3.

Freeman, J. B. (2018). Doing psychological science by hand [Publisher: SAGE Publications Inc]. *Current Directions in Psychological Science*, *27*(5), 315–323.

Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1436), 2435–2440. Retrieved September 13, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1690464/

Gagné, N., & Franzen, L. (2023). How to run behavioural experiments online: Best practice suggestions for cognitive psychology and neuroscience. *Swiss Psychology Open: the official journal of the Swiss Psychological Society*, *3*(1), 1.

Gallivan, J. P., Chapman, C. S., Wolpert, D. M., & Flanagan, J. R. (2018). Decision-making in sensorimotor control [Number: 9 Publisher: Nature Publishing Group]. *Nature Reviews Neuroscience*, *19*(9), 519–534.

Garaizar, P., & Reips, U.-D. (2019). Best practices: Two web-browser-based methods for stimulus presentation in behavioral experiments with high-resolution timing requirements. *Behavior Research Methods*, *51*(3), 1441–1453.

Greenaway, A.-M., Nasuto, S., Ho, A., & Hwang, F. (2021). Is home-based webcam eye-tracking with older adults living with and without alzheimer's disease feasible? *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, 1–3.

Grootswagers, T. (2020). A primer on running human behavioural experiments online. *Behavior Research Methods*, *52*(6), 2283–2286.

Gudi, A., Li, X., & van Gemert, J. (2020). Efficiency in real-time webcam gaze tracking. In A. Bartoli & A. Fusiello (Eds.), *Computer vision – ECCV 2020 workshops* (pp. 529–543). Springer International Publishing.

Hansen, D. W., & Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze [Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 478–500.

Harrison, P. M. c. (2020). psychTestR: An r package for designing and conducting behavioural psychological experiments. *Journal of Open Source Software*, *5*(49), 2088.

Hartshorne, J. K., de Leeuw, J. R., Goodman, N. D., Jennings, M., & O'Donnell, T. J. (2019). A thousand studies for the price of one: Accelerating psychological science with pushkin. *Behavior Research Methods*, *51*(4), 1782–1803.

Hauger, D., Paramythis, A., & Weibelzahl, S. (2011). Using browser interaction data to determine page reading behavior. In J. A. Konstan, R. Conejo, J. L. Marzo,

& N. Oliver (Eds.), *User modeling, adaption and personalization* (pp. 147–158). Springer.

Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, *7*(1), 43–64.

Heck, M., Becker, C., & Deutscher, V. (2023, January 3). *Webcam eye tracking for desktop and mobile devices: A systematic review*. Retrieved July 13, 2023, from https://hdl.handle.net/10125/103459

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, *18*(3), 384–401.

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). Lab.js: A free, open, online study builder. *Behavior Research Methods*, *54*(2), 556–573.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD [Number: 7302 Publisher: Nature Publishing Group]. *Nature*, *466*(7302), 29–29.

Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The 'real-world approach' and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, *11*, 721.

Hsu, K. J., Caffey, K., Pisner, D., Shumake, J., Risom, S., Ray, K. L., Smits, J. A. J., Schnyer, D. M., & Beevers, C. G. (2018). Attentional bias modification treatment for depression: Study protocol for a randomized controlled trial. *Contemporary Clinical Trials*, *75*, 59–66.

Hutt, S., Wong, A., Papoutsaki, A., Baker, R. S., Gold, J. I., & Mills, C. (2023). Webcam-based eye tracking to detect mind wandering and comprehension errors. *Behavior Research Methods*.

Jacobs, O. L. E., Pazhoohi, F., & Kingstone, A. (2023). Contrapposto posture captures visual attention: An online gaze tracking experiment [Publisher: Routledge _eprint: https://doi.org/10.1080/13506285.2023.2213904]. *Visual Cognition*, *31*(2), 160–167.

Johnson, B. P., Dayan, E., Censor, N., & Cohen, L. G. (2021). Crowdsourcing in cognitive and systems neuroscience [Publisher: SAGE Publications Inc STM]. *The Neuroscientist*, 10738584211017018.

Jun, E., Hsieh, G., & Reinecke, K. (2017). Types of motivation affect study selection, attention, and dropouts in online experiments. *Proceedings of the ACM on Human-Computer Interaction*, *1*, 56:1–56:15.

Kar, A., & Corcoran, P. (2016). Towards the development of a standardized performance evaluation framework for eye gaze estimation systems in consumer platforms. *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 002061–002066.

Khamis, M., Alt, F., Hassib, M., von Zezschwitz, E., Hasholzner, R., & Bulling, A. (2016). GazeTouchPass: Multimodal authentication using gaze and touch on mobile devices. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2156–2164.

Khamis, M., Hassib, M., Zezschwitz, E. v., Bulling, A., & Alt, F. (2017). GazeTouchPIN: Protecting sensitive data on mobile devices using secure multimodal authentication. *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 446–450.

Klaib, A. F., Alsrehin, N. O., Melhem, W. Y., & Bashtawi, H. O. (2019). IoT smart home using eye tracking and voice interfaces for elderly and special needs people. *Journal of Communications*, 614–621.

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, *67*(4), 151–185.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13852–13857.

Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, *25*(3), 296–324.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities. *Vision Research*, *41*(25), 3559–3565.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328.

Lavoie, E. B., Valevicius, A. M., Boser, Q. A., Kovic, O., Vette, A. H., Pilarski, P. M., Hebert, J. S., & Chapman, C. S. (2018). Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. *Journal of Vision*, *18*(6), 1–20.

Leeuw, J. R. d., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351.

Levine, J. L. (1981). *An eye-controlled computer*. IBM Research Division, TJ Watson Research Center.

Logitech 2023 annual report, SEC form 10-k (Annual Report). (2023). Logitech. Retrieved August 24, 2023, from https://s1.q4cdn.com/104539020/files/doc_financials/2023/ar/71e634f7-ce7a-4c9c-95cf-5a98d8fb9068.pdf

Lourenco, S. F., & Tasimi, A. (2020). No participant left behind: Conducting science during COVID-19. *Trends in Cognitive Sciences*, *24*(8), 583–584.

Lovett, M., Bajaba, S., Lovett, M., & Simmering, M. J. (2018). Data quality from crowdsourced surveys: A mixed method inquiry into perceptions of amazon's mechanical turk masters [Publisher: John Wiley & Sons, Ltd]. *Applied Psychology*, *67*(2), 339–366.

Lukács, G., & Gartus, A. (2022). Precise display time measurement in JavaScript for web-based experiments. *Behavior Research Methods*.

Ma, H. L., Bertrand, J. K., Chapman, C. S., & Hayward, D. A. (2023). You read my mind: Generating and minimizing intention uncertainty under different social contexts in a two-player online game [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: Human Perception and Performance*, No Pagination Specified–No Pagination Specified.

Madsen, J., Júlio, S. U., Gucik, P. J., Steinberg, R., & Parra, L. C. (2021). Synchronized eye movements predict test scores in online video education [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *118*(5), e2016980118.

Maldonado, M., Dunbar, E., & Chemla, E. (2019). Mouse tracking as a window into decision making. *Behavior Research Methods*, *51*(3), 1085–1101.

Mason, W., & Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavior Research Methods*, *44*(1), 1–23.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324.

Meng, C., & Zhao, X. (2017). Webcam-based eye movement analysis using CNN [Conference Name: IEEE Access]. *IEEE Access*, *5*, 19581–19587.

Miller, R., Schmidt, K., Kirschbaum, C., & Enge, S. (2018). Comparability, stability, and reliability of internet-based mental chronometry in domestic and laboratory settings. *Behavior Research Methods*, *50*(4), 1345–1358.

Moss, A., Rosenzweig, C., Jaffe, S. N., Gautam, R., Robinson, J., & Litman, L. (2021). *Bots or inattentive humans? identifying sources of low-quality data in online platform* (Type: article).

Moss, A. J., Rosenzweig, C., Robinson, J., Jaffe, S. N., & Litman, L. (2023). Is it ethical to use mechanical turk for behavioral research? relevant data from a representative survey of MTurk participants and wages. *Behavior Research Methods*.

Necka, E. A., Cacioppo, S., Norman, G. J., & Cacioppo, J. T. (2016). Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community participants [Publisher: Public Library of Science]. *PLOS ONE*, *11*(6), e0157732.

Newman, A., Bavik, Y. L., Mount, M., & Shao, B. (2021). Data collection via online platforms: Challenges and recommendations for future research. *Applied Psychology*, *70*(3), 1380–1402.

Ouellette Zuk, A. A., Bertrand, J. K., & Chapman, C. S. (2023, June 7). Continuous measures of decision-difficulty captured remotely: I. mouse-tracking sensitivity extends to tablets and smartphones [Pages: 2023.06.06.543796 Section: New Results].

Paolacci, G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, *5*(5), 9.

Papoutsaki, A., Laskey, J., & Huang, J. (2017). SearchGazer: Webcam eye tracking for remote studies of web search. *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, 17–26.

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions.

Patla, A., & Vickers, J. (2003). How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental Brain Research*, *148*(1), 133–138.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Pronk, T., Wiers, R. W., Molenkamp, B., & Murre, J. (2020). Mental chronometry in the pocket? timing accuracy of web applications on touchscreen and keyboard devices. *Behavior Research Methods*, *52*(3), 1371–1382.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search [Publisher: SAGE Publications]. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

Rayner, K., & Reingold, E. M. (2015). Evidence for direct cognitive control of fixation durations during reading. *Current Opinion in Behavioral Sciences*, *1*, 107–112.

Robal, T., Zhao, Y., Lofi, C., & Hauff, C. (2018). Webcam-based attention tracking in online learning: A feasibility study. *23rd International Conference on Intelligent User Interfaces*, 189–197.

Salminen, J., Jansen, B. J., An, J., Jung, S.-G., Nielsen, L., & Kwak, H. (2018). Fixation and confusion: Investigating eye-tracking participants' exposure to information in personas. *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, 110–119.

San Agustin, J., Skovsgaard, H., Mollenbach, E., Barret, M., Tall, M., Hansen, D. W., & Hansen, J. P. (2010). Evaluation of a low-cost open-source gaze tracker. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, 77–80.

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences*, *10*(4), 251.

Saxena, S., Lange, E., & Fink, L. (2022). Towards efficient calibration for webcam eye-tracking in online experiments. *2022 Symposium on Eye Tracking Research and Applications*, 1–7.

Schneegans, T., Bachman, M., Huettel, S., & Heekeren, H. (2021). *Exploring the potential of online webcam-based eye tracking in decision-making research and influence factors on data qualit* (Type: article).

Schröter, I., Grillo, N. R., Limpak, M. K., Mestiri, B., Osthold, B., Sebti, F., & Mergenthaler, M. (2021). Webcam eye tracking for monitoring visual attention in hypothetical online shopping tasks [Number: 19 Publisher: Multidisciplinary Digital Publishing Institute]. *Applied Sciences*, *11*(19), 9281.

Scott, K., & Schulz, L. (2017). Lookit (part 1): A new online platform for developmental research. *Open Mind*, *1*(1), 4–14.

Semmelmann, K., Hönekopp, A., & Weigelt, S. (2017). Looking tasks online: Utilizing webcams to collect video data from home. *Frontiers in Psychology*, *8*(1582), 1–11.

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451–465.

Sewell, W., & Komogortsev, O. (2010). Real-time eye gaze tracking with an unmodified commodity webcam employing a neural network. *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, 3739–3744.

Shehu, I. S., Wang, Y., Athuman, A. M., & Fu, X. (2021b). Remote eye gaze tracking research: A comparative evaluation on past and recent progress [Number: 24 Publisher: Multidisciplinary Digital Publishing Institute]. *Electronics*, *10*(24), 3165.

Shevchenko, Y. (2022). Open lab: A web application for running and sharing online experiments. *Behavior Research Methods*, *54*(6), 3118–3125.

Skovsgaard, H., Agustin, J. S., Johansen, S. A., Hansen, J. P., & Tall, M. (2011). Evaluation of a remote webcam-based eye tracker. *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications*, 1–4.

Slim, M. S., & Hartsuiker, R. J. (2022). Moving visual world experiments online? a web-based replication of dijkgraaf, hartsuiker, and duyck (2017) using PCIbex and WebGazer.js. *Behavior Research Methods*.

Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences*, *21*(10), 736–748.

Stillman, P. E., Krajbich, I., & Ferguson, M. J. (2020). Using dynamic monitoring of choices to predict and understand risk preferences [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *117*(50), 31738–31747.

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences*, *22*(6), 531–543.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments [Publisher: SAGE Publications Inc]. *Teaching of Psychology*, *44*(1), 24–31.

Stone, S. A., Boser, Q. A., Dawson, T. R., Vette, A. H., Hebert, J. S., Pilarski, P. M., & Chapman, C. S. (2022). Generating accurate 3d gaze vectors using synchronized eye tracking and motion capture. *Behavior Research Methods*.

Stone, S. A., & Chapman, C. S. (2023). Unconscious frustration: Dynamically assessing user experience using eye and mouse tracking. *Proceedings of the ACM on Human-Computer Interaction*, *7*, 168:1–168:17.

Thomas, K. A., & Clifford, S. (2017). Validity and mechanical turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, *77*, 184–197.

Uittenhove, K., Jeanneret, S., & Vergauwe, E. (2022). *From lab-testing to web-testing in cognitive research: Who you test is more important than how you tes* (Type: article).

Van Doorn, G., Woods, A., Levitan, C. A., Wan, X., Velasco, C., Bernal-Torres, C., & Spence, C. (2017). Does the shape of a cup influence coffee taste expectations? a cross-cultural, online study. *Food Quality and Preference*, *56*, 201–211.

Veselovsky, V., Ribeiro, M. H., & West, R. (2023, June 13). Artificial artificial artificial intelligence: Crowd workers widely use large language models for text production tasks. Retrieved July 27, 2023, from http://arxiv.org/abs/2306.07899

Vos, M., Minor, S., & Ramchand, G. C. (2022). Comparing infrared and webcam eye tracking in the visual world paradigm [Accepted: 2023-01-03T13:11:04Z Publisher: University of California Press].

Wadsworth, H. E., Galusha-Glasscock, J. M., Womack, K. B., Quiceno, M., Weiner, M. F., Hynan, L. S., Shore, J., & Cullum, C. M. (2016). Remote neuropsychological assessment in rural american indians with and without cognitive impairment. *Archives of Clinical Neuropsychology*, *31*(5), 420–425.

Walter, S. L., Seibert, S. E., Goering, D., & O'Boyle, E. H. (2019). A tale of two sample sources: Do results from online panel data and conventional data converge? *Journal of Business and Psychology*, *34*(4), 425–452.

Weydmann, G., Palmieri, I., Simões, R. A. G., Centurion Cabral, J. C., Eckhardt, J., Tavares, P., Moro, C., Alves, P., Buchmann, S., Schmidt, E., Friedman, R., & Bizarro, L. (2022). Switching to online: Testing the validity of supervised remote testing for online reinforcement learning experiments. *Behavior Research Methods*.

Wisiecka, K., Krejtz, K., Krejtz, I., Sromek, D., Cellary, A., Lewandowska, B., & Duchowski, A. (2022). Comparison of webcam and remote eye tracking. *2022 Symposium on Eye Tracking Research and Applications*, 1–7.

Wispinski, N. J., Gallivan, J. P., & Chapman, C. S. (2020). Models, movements, and minds: Bridging the gap between decision making and action. *Annals of the New York Academy of Sciences*, *1464*(1), 30–51.

Wong, A. Y., Bryck, R. L., Baker, R. S., Hutt, S., & Mills, C. (2023). Using a webcam based eye-tracker to understand students' thought patterns and reading behaviors in neurodivergent classrooms. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 453–463.

Xu, P., Ehinger, K. A., Zhang, Y., Finkelstein, A., Kulkarni, S. R., & Xiao, J. (2015, May 20). TurkerGaze: Crowdsourcing saliency with webcam based eye tracking. Retrieved August 24, 2022, from http://arxiv.org/abs/1504.06755

Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, *16*(6), 1485–1505. Retrieved August 24, 2022, from https://journal.sjdm.org/21/210525/jdm210525.html

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In A. L. Yarbus (Ed.), *Eye movements and vision* (pp. 171–211). Springer US.

Young, L. R., & Sheena, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, *7*(5), 397–429.

Zehr, J., & Schwarz, F. (2018). PennController for internet based experiments (IBEX) [Publisher: OSF].

Zhang, X., Sugano, Y., & Bulling, A. (2019). Evaluation of appearance-based methods and implications for gaze-based applications. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13.

Zhao, Y., Lofi, C., & Hauff, C. (2017). Scalable mind-wandering detection for MOOCs: A webcam-based approach. In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, & M. Pérez-Sanagustín (Eds.), *Data driven approaches in digital education* (pp. 330–344). Springer International Publishing.

Zheng, C., & Usagawa, T. (2018). A rapid webcam-based eye tracking method for human computer interaction [ISSN: 2475-7896]. *2018 International Conference on Control, Automation and Information Sciences (ICCAIS)*, 133–136.

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, *111*(4), 493–504.

Chapter 2

Assessing webcam eye-tracking utility in digital object interactions

Abstract

Do patterns of eye-hand coordination observed during real-world object interactions apply to digital, screen-based object interactions? We adapted a real-world object interaction task (physically transferring cups in sequence about a tabletop) into a two-dimensional screen-based task (dragging-and-dropping circles in sequence with a cursor). We collected gaze (with webcam eye-tracking) and cursor position data from 51 fully-remote, crowd-sourced participants who performed the task on their own computer. We applied real-world time-series data segmentation strategies to resolve the self-paced movement sequence into phases of object interaction and rigorously cleaned the webcam eye-tracking data. In this preliminary investigation, we found that: 1) real-world eye-hand coordination patterns persist and adapt in this digital context, and 2) remote, online, cursor-tracking and webcam eye-tracking are useful tools for capturing visuomotor behaviours during this ecologically-valid human-computer in-

A version of this work was previously published as: Bertrand, J. K. & Chapman, C. S. (2023). Dynamics of eye-hand coordination are flexibly preserved in eye-cursor coordination during an online, digital, object interaction task. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3544548.3580866.

teraction task. We discuss how these findings might inform design principles and further investigations into natural behaviours that persist in digital environments.

2.1 Introduction

At the core of all well-considered user experiences is the user themselves. So-called human-centered designs incorporate behaviour, cognition, and perception into their product. An early example is the psychophysical mapping of human sensitivity to flickering light. Over hundreds of years, scientists learned that a light display refreshing at a minimum of 30 Hz appeared continuous to the human eye - setting the benchmark for early computer screens.

Of course, perceiving the world is only part of a user's experience - they also interact with their environment to achieve their goals. Like the display refresh-rate example, principles of human interaction can dictate good design. Fitts' Law (Fitts, 1954) is one such finding that is now adopted as a design principle in human-computer interaction (HCI; MacKenzie, 1992; Seow, 2005). In this seminal work, Fitts (1954) quantified the speed-accuracy tradeoff for movements of different amplitudes (how far you need to move) to targets of different sizes. Put succinctly, he showed a lawful relationship whereby larger amplitude movements and smaller targets both result in longer movement times. These real-world findings have since been explored in depth in an HCI context (MacKenzie, 1992), informing and assessing the design of two- (MacKenzie & Buxton, 1992) and three-dimensional (Grossman & Balakrishnan, 2004) pointing devices, the soft (virtual) QWERTY keyboard (Mackenzie et al., 1999; William Soukoreff & Scott Mackenzie, 1995), and the properties and placement of interactive web elements (Karousos et al., 2013; Lin & Ho, 2020; McGuffin & Balakrishnan, 2005; Roy et al., 2021).

Critically, user experience (UX) is best thought of as a dynamic cycle of perception and action whereby the information we need to guide our upcoming actions is informed by where we look; how we move then shapes the environment causing changes in the perceptual experience. Consider, for example, the coordinated effort required of the visual and motor systems to safely pick up a cup full of hot coffee. Before any movement, the eyes will fixate the drink, leading the action of the hand by about 500 milliseconds. Seamlessly, as soon as the mug is grasped, the eyes will move to look at the next object for action, like a sugar packet, well before the hand is finished moving the hot drink (Land et al., 1999). These patterns of visuomotor coordination are ubiquitous and stereotypic, appearing in human and non-human primates (Arora, 2019; Ngo et al., 2022) alike. Here we ask, in the same way real-world findings have informed computer screen and keyboard design, can principles of real-world eye-hand coordination help inform UX design for digital interactions?

To approach this question, we used an online, eye-tracking-enabled platform (Labvanced; Finger et al., 2017) to create a screen-based version of a real-world object interaction task (Lavoie et al., 2018). Instead of moving cups to targets on a table, crowdsourced participants (N = 51) dragged circles to targets on their computer screen while we recorded cursor movements and webcam eye-gaze coordinates. We had two primary motivations - first, to explore if the quality of the webcam gaze data (and subsequent processing procedures) would be sufficient to explore visuomotor coordination in our specific task, and second, to quantify if the patterns of eye-cursor coordination would match principles of eye-hand coordination in the real world. Our findings, while preliminary, reveal that both of these are true for this experiment, offering an introduction to entirely new ways of collecting user experience data and suggesting UX design should further explore and consider the tight and environment-invariant principles of visuomotor coordination during target interaction tasks.

2.2 Related Works

2.2.1 Real-World Object Interactions

The tight coupling, in both space and time, of eye movements and motor actions has been well-documented. While much of this research has involved rigid, paradigmatic tasks, those most relevant to UX leverage technological advances to explore eye-hand coordination during self-paced, natural and realistic interactions. For example, the seminal works of Land et al. and Hayhoe measured where people look when making a pot of tea (Land et al., 1999) and preparing a sandwich (Hayhoe, 2000). Despite the complexity of these tasks and the lack of experimental structure, the researchers were able to break down the tasks into their constituent subtasks (e.g. reaching for the kettle, removing the lid, etc.) to reveal remarkably consistent patterns of eye-hand coordination (Land & Hayhoe, 2001).

First, even though there are irrelevant objects throughout the kitchen scenes, the eyes only ever fixate on task-relevant objects (Land & Hayhoe, 2001). Critically, these are not the most salient (as defined by low level visual properties like contrast) objects in the visual field; rather, gaze only lands upon task-relevant objects. Second, gaze behaves serially, always fixed upon the current object of manipulation, leading the hand to that object, and when the manipulation is almost complete, moving on to the next object without return (Land & Hayhoe, 2001). In Land et al.'s tea-making task (1999), aggregated across all 94 distinct sub-tasks, the dynamics of the gaze and hand display a consistent pattern: participants fixate on the object to be manipulated for about half a second prior to the hand's initial movement towards that object and remain fixated there until leaving for the next target about half a second before the completion of the current manipulation.

This general pattern of the eye leading motor action has been found in many contexts, extending beyond kitchen activities to less obvious forms of naturalistic, visually-guided interactions like walking (Land, 2006; Patla & Vickers, 2003), keyboard typing (Butsch, 1932), and music playing (Furneaux & Land, 1999). Across these interactions, the exact amount of time the eye leads the hand appears to be at least 500 ms (and up to about 1 second), showing some flexibility for the time (Deconinck et al., 2011) or accuracy constraints (Rand & Stelmach, 2010) of the task, or kinematics required for different task contexts (Johansson et al., 2001; Pelz et al., 2001). Lavoie et al. (2018) took a modern approach to investigations of eye-hand coordination during real-world object interaction by combining state of the art motion capture and mobile eye-tracking. In perfect alignment with Land et al. (1999), Lavoie et al. (2018) found that all object interactions involved the eyes fixating the object at least 500 ms prior to the start of the interaction. Then, within 600 ms of the interaction start, the eyes would leave to look ahead to the next area for interaction (Lavoie et al., 2018). This dominant pattern of eye-hand coordination has proven itself highly consistent across real-world object interactions, but, in service of applying real-world findings to digital domains, what happens when we move towards interactions with digital objects?

2.2.2 Lab-based Digital Interactions

Some visuomotor coordination research trades the complexity of an in-lab kitchen for the control of a computer workstation, treating screen-based, digitally-presented objects as a convenient proxy for real-world objects. Nonetheless, patterns of eye leading hand (or computer cursor or manipulandum) are remarkably consistent in the digital domain. In tracing shapes with a cursor, the eye leads the cursor by 223-295 ms (Deng et al., 2016), and in distractorless visual search there is a 190 ms lead (Bieg et al., 2010). A related paradigm of tracking an unpredictable object on a screen shows that the eye lags behind the target object by 24 ms, whereas the hand lags behind it by 108 ms (Danion & Flanagan, 2018). During simple reaches towards on-screen targets, the eyes arrive about 386 ms before the hand (Sailer et al., 2000), although gaze and hand dynamics are flexible to factors like the target's visibility before or during the reach (van Donkelaar & Staub, 2000). Adaptive gaze behaviour is also shown for two-target sequential reaches: the eye anchors to the first target for an extra 95 ms to ensure the hand's arrival before continuing to a second target (Rand & Stelmach, 2010). For visually-guided sequential movements of a manipulandum-handle's contact with 5 virtual target objects, the gaze arrives at targets 208 ms before contact and leaves 106 ms after contact (Bowman et al., 2009). Finally, dragging virtual objects about in a pseudo-touchscreen context also elicits gaze-leading-hand patterns (Sims et al., 2011). Taken together, this line of research predominantly studies target-directed reaching and following, but not interaction. Even so, in all contexts, the eyes lead the hand (or cursor) by a few hundred milliseconds.

2.2.3 UX-Focused Digital Interactions

What do eye-cursor coordination patterns look like when digital object interactions are also realistic and ecologically-valid? The HCI domain offers us some answers, however object interactions beyond simple target clicks remain almost entirely unexplored. Early eye-cursor studies employed search engine results pages (SERPs; Guo and Agichtein, 2010a; Navalpakkam et al., 2013; Rodden and Fu, 2007) but mostly considered visuomotor coordination only from a spatial context, measuring the pixel distance between the cursor and eye. Huang et al. (2012) did consider timing, finding that the gaze led the cursor by at least 250, and typically 700 ms during SERP browsing. Of course, with less experimental control, scientists are more likely to find a range of behaviours. Indeed, B. A. Smith et al. (2000) looked at cursor-pointing to graphical user interface targets and observed at least three eye-cursor coordination patterns: "eye gaze following the cursor to the target", "eye gaze leading the cursor to the target", and less commonly, "eye gaze switching between the cursor and target until the target is reached". The use of multiple strategies during digital yet ecologically-valid tasks is perhaps best illustrated by Liebling and Dumais (2014). Here subjects performed their regular work duties on their office desktop while their gaze and cursor movements were recorded. To begin making sense of the data, the researchers anchored their analysis to cursor clicks, with 32 classes of click-targets determined by metadata records. This rich dataset revealed nuanced coordination - in general, the gaze arrived near the click point 100-200 ms earlier than the cursor. However, the occurrence of or need for coordination varied by target-type: the gaze led the cursor ~30% of the time for 'TitleBar' clicks, yet 'List' clicks were gaze-first more than 85% of the time (Liebling & Dumais, 2014). In these unconstrained tasks we find additional evidence to support that the eyes lead the cursor, but also see implications for the limitation of this approach when faced with real digital interaction complexity. In the current study, we attempt to strike a balance between ecological validity and experimental control, focusing on a prescribed digital interaction sequence without imposing any constraints on how (e.g. where to look, how fast to move etc.) the sequence should be completed.

2.2.4 Webcam Eye-tracking

One limitation of almost all of the aforementioned studies is that they occur in the lab. A principle of human-centered design is not only to focus on the user but also to consider the environment and context in which their experience is happening (ISO 9241-210:2010). One recent technological advancement that might make it possible to study visuomotor coordination in more authentic environments (e.g. users in their own homes on devices they regularly use) is to use webcam data to derive estimates of screen-based gaze behaviour. However, webcam eye-tracking has struggled to establish its utility as a research tool. The reasons are numerous: webcam eye-tracking has a much slower sampling rate (~10 Hz compared with 100+ Hz in the lab; Bánki et al., 2022; Gagné and Franzen, 2023; Semmelmann and Weigelt, 2018), many users are unable to participate due to a slow internet connection or insufficient hardware (Bánki et al., 2022; Gagné & Franzen, 2023), uncontrolled lighting can significantly decrease data quality (Fraser et al., 2021; Semmelmann & Weigelt, 2018; Yang &

Krajbich, 2021) and, even with optimal conditions, extensive time must be spent calibrating the system (up to 50% of the study duration as in Semmelmann and Weigelt (2018)). Despite these limitations, recent advances, especially in using machine learning to predict gaze location (e.g. Labvanced v2 High Sampling Mode eye-tracking; Finger et al., 2017), offer a path forward, especially where spatial accuracy is the most important feature of the data (Semmelmann & Weigelt, 2018; Wisiecka et al., 2022). Further, researchers have shown that focusing on fixations to the most relevant areas (i.e. areas of interest, or AOIs) is a reasonable approach for eye-tracking data (Holmqvist et al., 2011). Therefore, a major motivation of the current study was to explore whether state-of-the-art webcam eye-tracking algorithms (Finger et al., 2017) combined with participation criteria (e.g. processing speed) and AOI-based clustering and analyses would provide sufficient data quality to explore eye-cursor coordination.

2.3 Methods

2.3.1 Participants

51 adults provided their informed consent to participate in the experiment. Of these, 14 participant datasets were rejected for unsalvageable eye data, and 8 participant datasets were rejected for low trial count (<50%) after removing trials with procedural or technical errors (see subsection 2.3.6 - Data Processing and Appendix B - Supplementary Materials for complete data cleaning procedure). The remaining 29 participants (12 female, 1 undisclosed gender; Age: $M = 27.07$, $SD = 10.75$) were 26 right-hand users and 3 left-hand users. All participants had no prior knowledge about the experiment or its objective. All experimental proceedings were approved by the University of Alberta's Research Ethics Board (Pro00087329) and were performed in accordance with relevant guidelines and regulations. All participants were recruited using the online crowdsourcing platform Prolific (www.prolific.co) and were paid for their time (6 GBP per hour, ~$10 CAD per hour).

2.3.2 Materials

All participant data was collected online using Labvanced (Finger et al., 2017), a browser-based Javascript experimentation platform. The Labvanced platform offers built-in webcam eye-tracking (Labvanced v2 High Sampling Mode eye-tracking Finger et al., 2017) and can record the position of the cursor across time. It was necessary to impose some minimum requirements to achieve stable data collection: only laptop (n = 19) or desktop (n = 10) computers with an audio output (headphones or speakers); only Mac (n = 3), Windows (n = 26) or Linux (n = 0) operating systems and Chrome browser; a webcam with a minimum resolution of 1280 x 720 pixels; a landscape-oriented screen with a minimum of 600 x 600 pixels ($Mode$ = 1920 x 1080 px) and a system (including internet connection) capable of collecting at least 10 samples per second of the head's position for optimal eye-tracking precision (M = 14.5 Hz, SD = 4.3 Hz).

2.3.3 Digital Task Layout

We modeled our digital task layout (see Figure 2.1A) to mirror Lavoie et al.'s 2018 Cups Task apparatus (see Figure 2.1B), which featured a short-walled table-top surface with a midline partition, two cups, 4 AOIs, a Home area, and a fixation sphere. In the real world, participants stood next to the counter-height apparatus, looking down on the surface. Thus, we designed our screen-based version to appear like a flat, bird's eye view of the real-world task. All real-world elements were proportionally scaled to a 800 x 450 pixel frame in Labvanced (and later, automatically scaled by Labvanced to each participant's screen resolution). Like the real-world version, the Far Left and Right AOIs (FLAOI/FRAOI) were colored blue, the Near Left and Right AOIs (NLAOI/NRAOI) were green, and the Home area was purple. The real-world white paper cups filled with white beads were modeled as white circles. As a proxy for haptic feedback, we designed the circles and Home area to be responsive to cursor hover by darkening in color whenever the cursor landed within their borders. Finally,

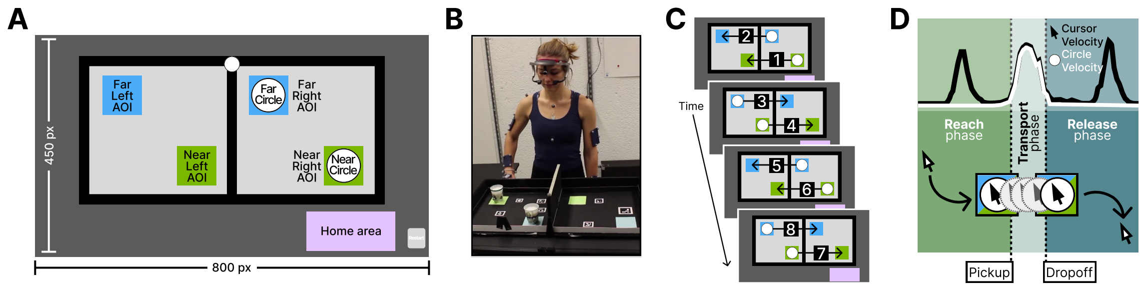


Figure 2.1: A) The digital task layout. The full screen (800 x 450 pixels) is depicted, where a purple Home area and a Restart button are located in the bottom right corner outside the black-bordered task area. The target Areas of Interest (AOIs) are split into those at the top (blue, Far Left and Right) and those at the bottom (green, Near Left and Right). Participants dragged-and-dropped circles (white, one Near, one Far) between target AOIs. B) A still image of the real-world Cup-Transfer Task we based our digital object interaction task on (from Lavoie et al., 2018). C) The task sequence. A trial involved 8 Moves, with the order and direction of circle movements shown with numbered arrows. The cursor started in the Home area to begin the trial and returned to the Home area after Moves 2, 4, 6, and 8. D) The segmentation of a single object interaction (i.e. one Move) into its events and phases. The top of the panel shows the velocity traces of the circle (white) and cursor (black). The onset of the circle movement was detected as a combination of these velocities and the distance between the cursor and the relevant target AOIs (bottom panel). Together, this defined the Pick-up and Drop-off events, which in turn defined the Reach (green), Transport (light blue) and Release (dark blue) phases.

select elements were introduced to the digital task space in service of loosely replicating the real-world, experimenter-guided experience: a restart button was always available should the participant realize they made a movement sequence error, text appeared with instructions if the participant took more than three seconds to move into their starting position, and a highlighted border appeared around the Home area to mark the important start and end events of the trial.

2.3.4 Task

Our task was an adaptation and extension of an established object interaction task from the real world (Lavoie et al., 2018). We doubled the number of object interactions within a sequence, and transformed the task to a digital, screen-based version. Critically, Lavoie's real-world task was designed to be segmented in time and space to allow for an examination of eye-hand coordination measures around the critical Pick-up and Drop-off interaction events (Lavoie et al., 2018). By adopting a similar structure, our analysis also relies on segmentation centered on these key time points. Our 8-movement sequence (see Figure 2.1C) is as follows: with the cursor always beginning at Home, Move 1 was a Pick-up of the Near circle at NRAOI, with its Drop-off at NLAOI. Immediately after, the Far circle was picked up from FRAOI and was transported to its Drop-off at FLAOI (Move 2). The cursor then returned to the Home position (as was required after every two object interactions). Moves 3 and 4 were a reflection of the first movements: the Far circle was picked up from FLAOI and moved back to FRAOI, and the Near Circle was picked up from NLAOI and moved back to NRAOI. After the cursor returned to Home, the Far circle was transported from FRAOI to FLAOI (Move 5), and then the Near circle moved from NRAOI to NLAOI (Move 6). This pattern was again reflected after the cursor visited Home, with Move 7 the pickup of the Near circle from NLAOI to drop off at NRAOI, and Move 8 the pickup of the Far circle from FLAOI to FRAOI. The cursor returned to Home to end the trial.

2.3.5 Procedure

Prolific (www.prolific.co) participants were provided a study link and a detailed study description that included an estimate of the study's duration (1 hour), the hardware requirements, and instructions for avoiding technical complications (included in Appendix B - Supplementary Materials). Clicking the study link launched the full-screen Labvanced window and requested webcam device permission. Participants failing to meet the minimum requirements would receive an error or warning message immediately. Barring no issues, participants would first read a consent form and provided they gave their informed consent, would then answer a brief demographic and hardware survey.

Next, participants would proceed self-paced through the task instructions. Following online research best practices (e.g. Gagné and Franzen, 2023), we developed extensive instructions including an instructional video (see Appendix B - Supplementary Materials), and gave task directions in a way that required participant engagement. Participants were informed that the task was a screen-based version of a real-world task. They were shown a picture of the real-world task (see Figure 2.1B) and told that the circles in their task were to be thought of as two-dimensional cups. Finally, participants were encouraged to use favourable lighting conditions and were instructed about the use of a virtual chinrest feature, strategies previously shown to improve webcam eye-tracking data quality (Semmelmann & Weigelt, 2018).

A 5 minute Labvanced eye-tracking calibration followed the instructions, and participants were required to repeat the calibration if the predicted gaze error exceeded 7% of the screen size. Lastly, participants completed one guided (click-through) practice trial, and then a second unguided practice trial with time-delayed hints (i.e. only shown if participant paused for three seconds or longer). Participants could repeat the unguided practice trial as many times as they wanted to ensure they understood the prescribed sequence of movements (1.2 unguided practice trials completed on average).

Participants would then complete the 50 self-paced experimental trials. Every 10 trials, they would receive an update about how many trials they had completed. A brief, 7-point eye-tracking re-calibration was performed every 5 trials, enabling the use of Labvanced's adaptive drift correction feature. After completing the 50 trials, participants were offered a long-form text input field to provide study feedback (if any) and thanked for their time. The study then concluded, with the browser exiting fullscreen mode, and participants receiving compensation via Prolific.

The entire experimental procedure, as a Labvanced study, can be accessed via the link in Appendix B - Supplementary Materials.

2.3.6 Data Processing

Employing webcam eye-tracking during an online, self-paced, sequential object-interaction task proved to be challenging. The resulting raw data required a number of quality assessments and treatments to ensure its utility for analysis. While this paper centers on our empirical findings, the corresponding methodological contribution of this work is not trivial, and we provide a detailed account of our data processing pipeline in Appendix B - Supplementary Materials.

The uncontrolled nature of the online testing environment could give rise to less accurate or spurious gaze predictions. We determined, in a cursory visual inspection of pilot data, that unlike cursor data, gaze data were prone to spatial distortions. That is, while much of the structure of the screen layout of the task was evident in most participants' gaze data (e.g. many fixations following a pattern shaped like the distribution of targets) these fixation "hot spots" would not necessarily project to the actual target locations - instead they were often shifted or skewed (see Figure 2.2 and Appendix B - Supplementary Materials for examples). However, if one is primarily interested in which object a person is fixating and when, the exact location of that fixation is mostly irrelevant, and instead you can define and analyze looks to AOIs in relative space. Taking advantage of the fact that our key analyses related to 4 distinct, spatially distributed locations for Pick-up and Drop-off events, we used a data-driven AOI approach. Here, we assumed that participants' gaze would primarily be driven toward the 4 target locations (NRAOI, NLAOI, FRAOI and FLAOI). Using data from all 50 trials, from all times when the participant had clicked and held the cursor button down, we employed a k-means clustering approach to spatially bin the gaze data into 4 corresponding AOIs (see Figure 2.2 for a representative participant's clustering centroids and see Appendix B - Supplementary Materials for additional examples). Thus, our eye-tracking data, while retaining its temporal resolution, was spatially transformed from the 800 x 450 Labvanced coordinate frame into four mutually-exclusive bins: NRAOI, NLAOI, FRAOI and FLAOI. Fourteen of the original 51 subject datasets were rejected because the clustering centroids did not follow the spatial configuration of the AOIs (left targets to the left of right targets, near targets lower on the screen than far targets), suggesting raw gaze data errors beyond a spatial distortion that we could not account for. In the Supplementary Materials (see Appendix B), we include a probability density analysis of the accepted clusters in transformed space where we fit bivariate normal distributions to each cluster for each participant and demonstrate that, on average, 28.97% of eye-tracking data, if linearly transformed to the Labvanced coordinate frame, would fall within the 80 x 80 pixel AOI it was cluster-assigned to. Importantly, only 0.1% of the eye data risked assignment to any of the 3 non-assigned AOIs in transformed pixel space.

Notably, this approach is not without its risks or limitations. First, by using the data to define the AOIs used for analysis, we run the risk of circularity. Therefore our first test was to ensure that the distribution of looks to each AOI across time matched the time-varying demands of the task. Since our clustering was collapsed across time, this would ensure that the reported looking behaviour was sensitive to the actual task being performed. Second, as discussed in Sections 2.4 - Results and 2.6 - Limitations and Future Directions, by only creating four cluster-based AOIs, we lose the ability to detect looks to other areas of the display (e.g. the Home or Fixation targets). Since these other areas were not relevant for the majority of our task, and entirely irrelevant to the key interaction events, this was a trade-off we felt was worth making despite the consequence of limiting our approach's applicability to other task designs or research questions. Third, the spatial accuracy assessment of our clustering approach (in Appendix B - Supplementary Materials) highlights the highly effective discrimination *between* AOIs by having effectively no chance of a mis-classified eye-gaze, but it also exposes the challenge of noisier eye-data for *within*-AOI discrimination - had we constrained ourselves to transformed eye data that fell within the boundaries of the actual on-screen AOI-objects, we would have lost more than

70% of the data. In Supplementary Materials (see Appendix B), we represent our clusters as independent bivariate normal probability density functions to visualize their clear spatial cohesiveness, but we acknowledge that the AOI-binning approach may have limited use where targets are less spatially distributed or the task space necessitates unpredictable, dynamic, numerous and/or densely organized targets.

Beyond the eye data cleaning, various steps (as outlined in Appendix B - Supplementary Materials) were taken to ensure participants completed the task correctly. While participants could move the circles about as they pleased, a trial was designed to only advance once all the movements were made. Although various real-time checks were performed to preemptively avoid sequence errors, they still occurred and those trials were removed from the analyzed data. Eight further subjects were removed for having a trial count below 50% after trial rejection for sequence (and other) errors. Therefore, 29 datasets were included in the following analyses.

2.3.7 Segmentation

In order to explore eye-cursor coordination patterns during object interactions, we needed to define and then automatically identify the 8 movements in each trial and the 2 object interactions (Pick-up and Drop-off) within each movement (see Figure 2.1D). This first necessitated re-sampling the cursor and eye data to a common sampling frequency of 60 Hz. Most often this meant that the eye data was upsampled while the cursor data was downsampled. Following Lavoie et al.'s 2018 approach, we considered the object interaction to include the period when the cursor moves toward the object (Reach: onset = cursor approaching Pick-up location + velocity exceed threshold; offset = Transport onset), the period when the cursor drags the object (Transport: onset = object leaving Pick-up location + velocity exceeds threshold; offset = object approaching Drop-off location + velocity drops below threshold) and the period when the cursor moves away from the object (Release: onset = Transport offset; offset = cursor leaving Drop-off location + velocity drops below threshold), as depicted in

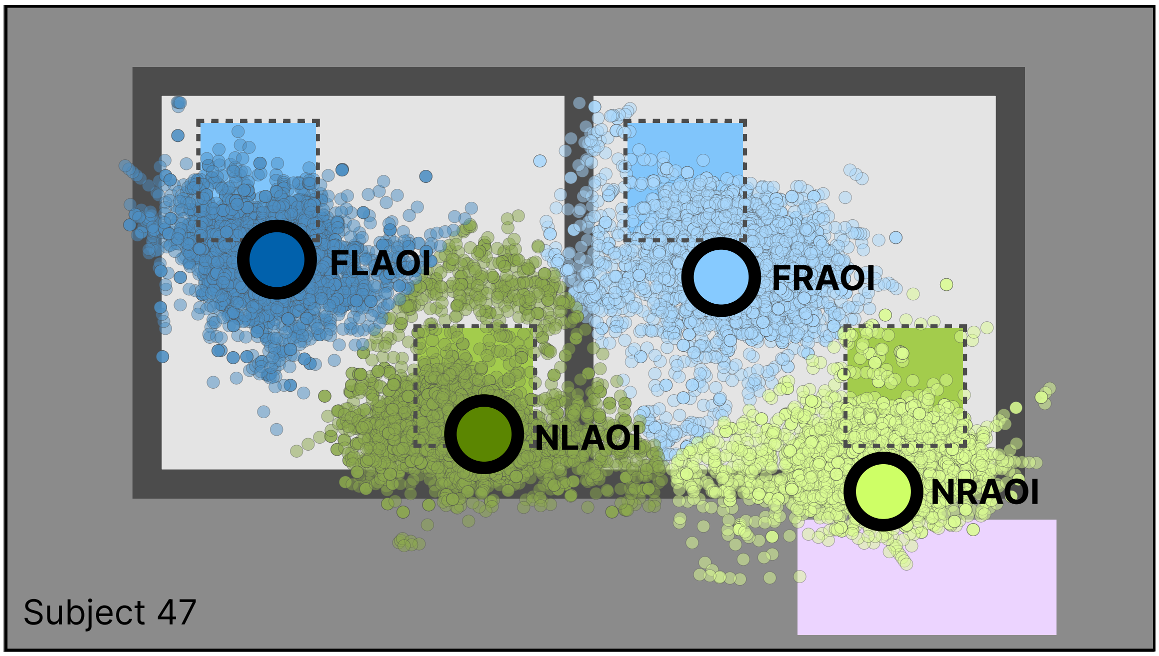


Figure 2.2: Mapping of one representative participant's raw eye-data to target AOIs. Raw eye-data samples (small circles) were clustered into 4 bins and, based on their spatial configuration, bins were assigned to one of the 4 target AOIs (dashed boxes). Individual samples are color-coded by their assigned bin: light green -> NRAOI, dark green -> NLAOI, light blue -> FRAOI and dark blue -> FLAOI. Raw-data bin-cluster centroids are shown in corresponding colors as filled circles with black borders. The raw data clearly groups into four clusters but the cluster centroids do not align with the actual task space (skewed down). Our analysis therefore relies on when the eye-data was within each cluster, not the actual space it occupied.

Figure 2.1D. A Pick-up is said to occur at the transition between Reach and Transport while a Drop-off is said to occur at the transition between Transport and Release. We used our custom Gaze and Movement Analysis (GaMA) software in MATLAB to segment our trial data into 8 movements (Moves 1-8) and each movement into the 3 phases (Reach, Transport and Release). In GaMA, spatial and temporal features of the data are used to automatically define key events, like the start of an object Transport (see Williams et al., 2019). We followed the principles of Lavoie et al. (2018) and other real-world examples (Valevicius et al., 2018; Williams et al., 2019) to segment our data, where thresholds were applied to the cursor and circles' velocity and AOI-proximity as a means to define the onset and offset of the Reach, Transport and Release phases (see Appendix B - Supplementary Materials for additional details and threshold values).

2.3.8 Data Analysis

Dependent Measures

We had two primary motivations: 1) to determine if the quality of webcam eye-tracking data would be sufficient to explore eye-cursor coordination dynamics for our specific 2D UX context and, if so, 2) to explore, in a preliminary way, if the $\sim$500 ms of fixation time around a manual interaction would be preserved in our specific 2D UX context, despite drastic differences in the physics and style of control.

With respect to 1), it is a hallmark of eye-hand coordination during object interaction that, even though not directly instructed, participants look almost exclusively at task relevant targets (namely the object they are going to interact with and the locations where they are going to move it to and from). As such, our first dependent measure examines, for each target location in the task (4 total, see subsection 2.3.3 - Digital Task Layout), the average time spent fixating that location when it was relevant to the current movement (i.e. a Pick-up or Drop-off location) and the average time spent fixating that location when it was irrelevant to the current movement (i.e. one of the two targets on every movement that are not a Pick-up or Drop-off location).

Average Fixation Duration (ms) Across a given trial, the average time spent fixating on one of the AOIs (NRAOI, NLAOI, FRAOI or FLAOI, see Figure 2.1A) when that AOI was a relevant location (a Pick-up or Drop-off location for the current movement) or not. Across the 8-movement sequence, each AOI was relevant and irrelevant an equal number of times.

With respect to 2), fixation time around an interaction consists of two values - how long the eyes are on an object prior to interaction (eye-arrival latency) and how long the eyes linger on an object after an interaction is initiated (eye-leaving latency). Our task involves object manipulations with two interaction events, the Pick-up and the Drop-off. Therefore, we examine the eye arrival and eye leaving latencies for both of these events.

Importantly, it has previously been reported that eye arrival and leaving latencies are not absolute (Lavoie et al., 2018), but can flexibly change based on the demands of the task in general and the durations of each constituent movement and phase in specific. Therefore, to test for these possible within-task adaptations, we also examined eye latencies and each movement in terms of the durations of the Reach, Transport and Release phases.

Based on this motivation, we extracted and analyzed the following measures per movement:

Phase Duration (ms) The time spent in each phase (Reach, Transport, Release).

Eye-arrival latency at Pick-up and Drop-off Eye-arrival latency (EAL) at Pick-up was defined as the difference between Transport start time and the time of the eye's arrival at the Pick-up location. EAL at Drop-off was defined as the difference between Transport end time and the time of the eye's arrival at the Drop-off location.

Eye-leaving latency at Pick-up and Drop-off Eye-leaving latency (ELL) at Pick-up was defined as the difference between Transport start time and the time of the eye leaving the Pick-up location. ELL at Drop-off was defined as the difference between Transport end time and the time of the eye leaving the Drop-off location.

Statistical Procedure

Each dependent measure was analyzed in Jamovi (Version 2.2.5; an open-source statistical software) using a two-factor repeated-measure analysis of variance (RMANOVA). If a two-way interaction was revealed from the omnibus RMANOVA, follow-up single-factor RMANOVAs were performed to test the simple main effects of one factor at all levels of the other factor. Significant main effects were explored with all pairwise comparisons. All reported RMANOVA p-values include a Greenhouse-Geisser correction for violations of sphericity, and all follow-up pairwise comparisons are fully reported in Appendix B - Supplementary Materials (with Bonferroni-corrected p-values).

2.4 Results

2.4.1 Online, webcam eye-tracking can be a suitable method for quantifying gaze behaviours

As described earlier, our data-driven definition of AOIs leaves us vulnerable to circularity in our analyses. To address this potential criticism, here we look at task relevant timing to check if our approach is valid. Since our AOI clustering is agnostic to timing, any effects of spatial gaze distribution across time that match task demands provide solid evidence for the utility of our approach. Critically, therefore, we show that participants' gaze fixated more on the Task Relevant AOIs than on the Task Irrelevant AOIs (Figure 2.3 - see Appendix B - Supplementary Materials for a complementary spatial analysis). These results align favourably with the early, real-world work of Hayhoe and Land (2000, 2006, 2001) and give credence to our use of webcam eye-tracking as a method for a preliminary investigation of gaze be-

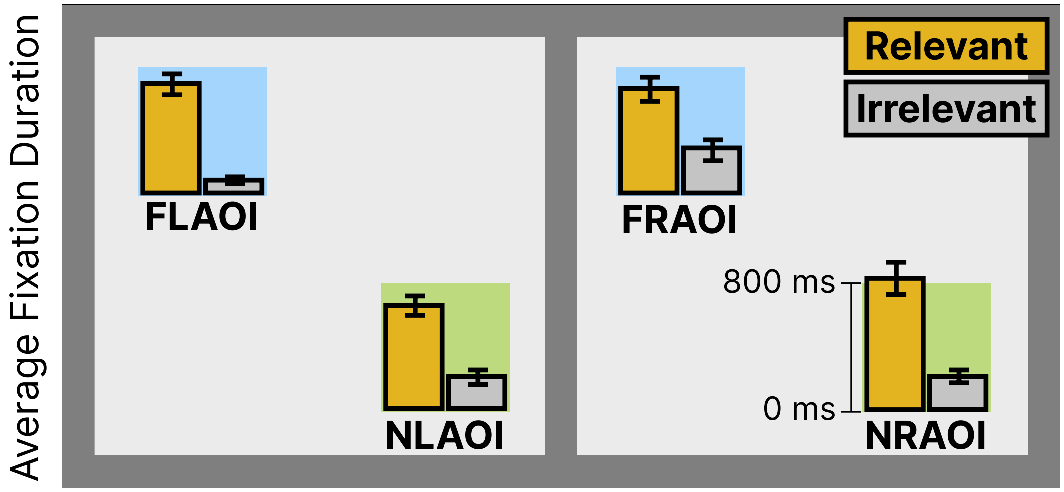


Figure 2.3: Average fixation duration to each of the four target AOIs (green and blue squares) across movements when that AOI was task-relevant (yellow bars, when that AOI was a Pick-up or Drop-off site for a movement) or task-irrelevant (grey bars). Relevant AOIs were fixated significantly longer than Irrelevant AOIs. Error bars show 95% confidence intervals of the estimated marginal means.

haviours during our specific online, screen-based, object interaction task. Our 4 x 2 (Position x Task Relevance [Relevant / Irrelevant]) RMANOVA revealed significant main effects of both Position ($F(1.50,\ 41.97) = 11.6$, $p < .001$) and Task Relevance ($F(1.00,\ 28.00) = 329.2$, $p < .001$), and a significant interaction between the two factors ($F(1.83{,}51.26) = 45.7$, p $<$ .001). Post-hoc pairwise follow-ups compared Relevant vs Irrelevant fixation durations at each location - for each location it was fixated more when it was Relevant than when it was Irrelevant (all p's $< .001$).

This analysis also allowed us to explore for any specific spatial biases in the eye-tracking data recorded in this task. In general, looks to targets were relatively evenly distributed, except for times when looks to other objects in the environment were mis-classified to spatially-proximal target locations. Specifically, at some times, looks to the Home position may have been categorized as looks to the NRAOI and looks to the Fixation position may have been categorized as looks to the FRAOI. This pattern is also visible in the complementary spatial analysis in Supplementary Materials (see

Appendix B), where the bivariate normal probability density function of the NRAOI shows more dispersion, likely as a result of also capturing some Home position looks. Along the same lines, in this supplemental analysis the FLAOI has the least dispersion, matching its status as the most isolated task-relevant object. As mentioned, however, the overall lack of spatial specificity is a consequence of our AOI clustering but does not appear to add significant noise to our analyses. For a complete analysis of these spatial biases, see Appendix B - Supplementary Materials.

In general, while accounting for the inherent limitations in the design of our study, this analysis offers a demonstration of the sensitivity of webcam eye-tracking. The stark differences in looking time driven by the expected pattern of task relevance, complemented by the low risk of assigning gazes to inaccurate clusters (as evidenced in Appendix B - Supplementary Materials), gave us sufficient confidence to further explore the dynamics of eye-cursor coordination.

2.4.2 Digital object interactions yield unique, context-specific movement dynamics

As described above, in order to understand the nuances of eye-cursor coordination it is first essential to map the naturally-occurring variations in task demand as indicated by the time spent in each movement and each phase within that movement (Reach, Transport and Release, Figure 2.4). Thus, we used a 3 x 8 (Phase x Movement) RMANOVA to examine phase duration. Both main effects of Phase ($F(1.58, 44.20) = 138.5$, $p < .001$) and Movement ($F(3.63, 101.68) = 34.1$, $p < .001$) were significant, as was their interaction ($F(5.38,150.53) = 21.0$, $p < .001$). Because we were most interested in learning how changes in each phase might impact eye latencies, we examined how phase values changed across movements. The three follow-up single-factor RMANOVAs (Reach/Transport/Release x Movement) each revealed main effects of Movement (Reach: $F(3.74, 104.80) = 24.5$, $p < .001$; Transport: $F(3.80, 106.41) = 22.2$, $p < .001$; Release: $F(3.13, 87.55) = 29.7$, $p < .001$). These results highlight a general pattern of longer Reach phases for movements covering longer screen distances. That is, movements that directly follow a Home visit (i.e. Moves 1, 3, 5, 7) cover more screen distance and elicit longer Reach phases than those that immediately follow a circle movement (i.e. Moves 2, 4, 6, 8) except for Move 1, which has the shortest Home to AOI distance. This finding aligns with the principles of Fitts' Law.

The single biggest difference between the movement dynamics in the digital compared to real-world task is the duration of the Transport phases. Cursor click and drag movements are *much* faster (around 200 ms) than their real-world counterparts (over 1000 ms from Lavoie et al., 2018). Besides a quick final move (M = 0.143 secs) and some slight differences between other movements, the Transport phases are relatively similar in duration (M's range from 0.205 to 0.255 secs, see Appendix B - Supplementary Materials for full pairwise analysis). Overall, the relative consistency of the Transport duration across the task again reflects that phase timing is primarily related to movement distance - Transport distance is the same for all movements.

Pairwise comparisons (as reported in Appendix B - Supplementary Materials) between movements for the Release phase also follow a pattern of longer phase durations for movements covering more screen distance (between the movement's drop-off location and the next destination location), as predicted by Fitts' Law.

2.4.3 Eye-cursor coordination during Pick-up interactions resembles the real world, while Drop-off coordination flexibly conforms to the digital context

Our final motivation was to understand eye-cursor coordination during digital object interaction by examining the latencies between the eye and cursor arriving (EAL) and leaving (ELL) the Pick-up and Drop-off sites across the 8 Movements (Figure 2.4). For each of EAL and ELL we ran an 8 x 2 (Movement x Interaction Site [Pick-up / Drop-off]) RMANOVA.

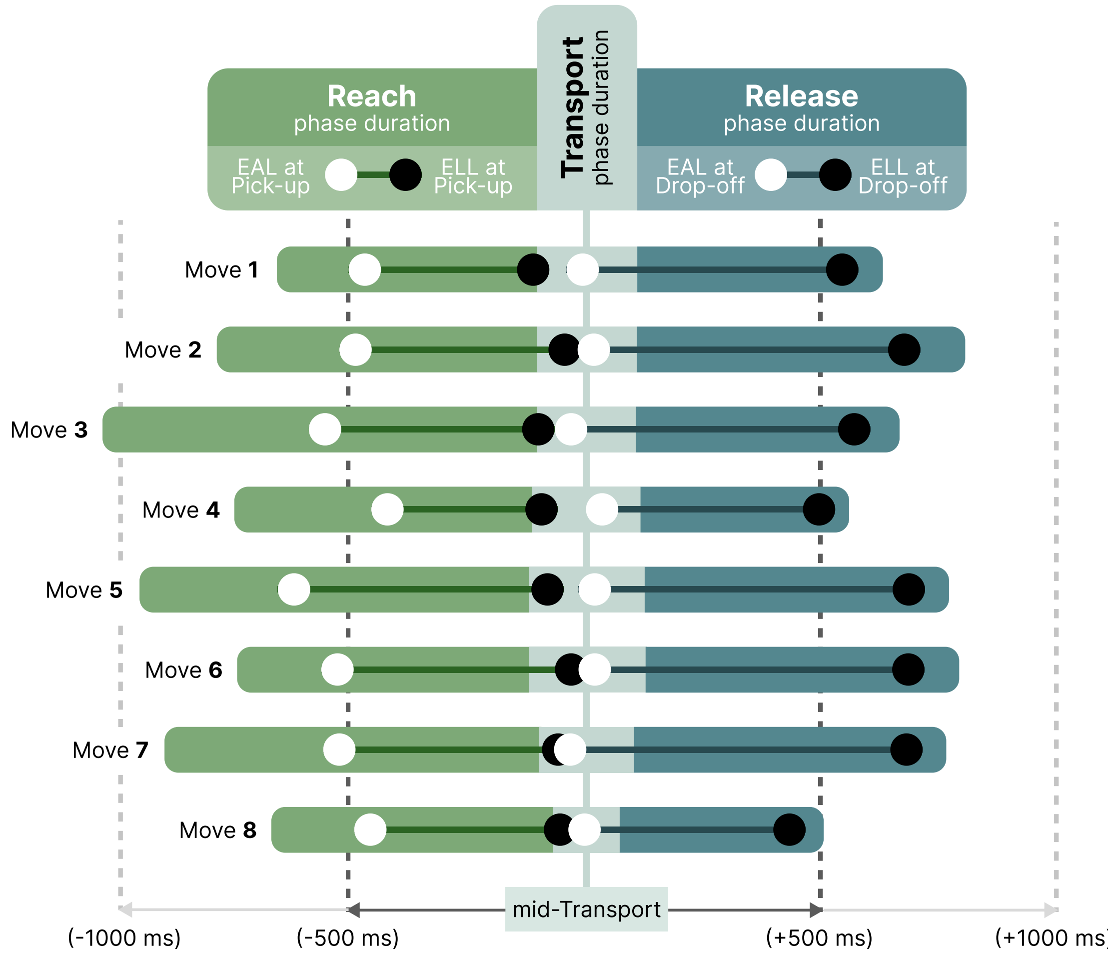


Figure 2.4: Interaction Phase durations and Eye Latencies at Pick-up and Drop-off across the 8 movements. Each movement is shown as a horizontal bar centered on the middle of the Transport (light blue) phase. Movements start with a Reach (green) as the cursor moves toward the target circle, transition to a Transport as the circle is moved, and end with a Release (dark blue) as the cursor moves away from the circle. The start of Transport is the Pick-up event and the end of Transport is the Drop-off event. White circles show the time the eye arrives (EAL) before each event while black circles show the time the eye leaves (ELL) following an event. Dark green lines connect the Pick-up EAL and ELL while Dark blue lines connect the Drop-off EAL and ELL. Despite the short Transport phases, the eye latencies adapt to ensure at least 500 ms of fixation time around each interaction event.

For EAL there were significant main effects of Movement ($F(3.43, 96.09) = 14.28$, $p < .001$) and Interaction Site ($F(1, 28) = 251.6$, $p < .001$), as well as a significant interaction between the two factors ($F(3.97, 111.2) = 8.37$, $p < .001$). Follow-up simple main effect RMANOVAs compared Pick-up and Drop-off EALs across the 8 movements. Drop-off EALs were remarkably consistent, showing no effect of Movement, demonstrating that gaze consistently arrives at a Drop-off location just over 100 ms before the clicked-and-dragged object. Pick-up EALs did show a significant effect of Movement ($F(3.87, 108.42) = 27.3$, $p = < .001$) which aligns with the duration of the Reach phase in which the Pick-up occurred. That is, for movements with a longer Reach phase (e.g. Movements 3, 5, 7) the eye arrives at the Pick-up location earlier - this kind of within-trial flexibility is also observed in the real world (Lavoie et al., 2018). Full reporting of the pairwise comparisons is available in Appendix B - Supplementary Materials.

Together, these EAL findings suggest: 1) similar to real-world interactions, during digital Pick-up interactions the eyes arrive about 400-500 ms before the cursor starts to move the object, and 2) unlike real-world interactions, during digital Drop-off interactions the eyes only arrive about 100-200 ms before the dragged object, reflecting the stark differences in the duration of digital versus physical object Transport.

For ELL there were significant main effects for both Movement ($F(2.93, 82.01) = 22.7$, $p < .001$) and Interaction Site ($F(1, 28) = 340.2$, $p < .001$), and also a significant two-way interaction ($F(4.62, 129.38) = 15$, $p < .001$). Follow-up simple main effect RMANOVAs compared Pick-up and Drop-off ELLs across the 8 movements and both were significant (Pick-up: $F(2.88, 80.7) = 8.2$, $p < .001$; Drop-off: $F(3.6, 100.91) = 24.3$, $p < .001$). Despite statistical differences, the timing of the eye leaving the Pick-up location is quite stable and short, ranging from $\sim$35-135 ms. Where Pick-up ELL does vary, it appears to change as a function both of the length of the upcoming transport and as a push-pull with the preceding eye arrival latencies. As an example, Movement 6 has a relatively long Transport phase and comparatively short preceding

EAL - this results in it having the longest Pick-up ELL. Full reporting of the pairwise comparisons is available in Appendix B - Supplementary Materials.

The most surprising result from our study, and what stands out as the biggest difference from real-world eye-hand coordination, is how long our participants spend looking at an object *after* they have dropped it off. Here, ELL at Drop-off exceeds 400 ms in all cases and is often more than 600 ms. This is drastically different from the real-world Drop-off ELLs which only range from 140-250 ms (Lavoie et al., 2018). This important finding demonstrates that participants compensate for abbreviated digital Transports by having their eyes remain fixated for longer at the location where the object is dragged to. This prolonged Drop-off ELL also scales with duration of the Release phase, with Movements with longer Release phases also showing the longest ELLs. This relationship suggests that the compensatory prolongation of the Drop-off ELL may in part relate to the planning of the next movement following a Release. Full reporting of the pairwise comparisons is available in Appendix B - Supplementary Materials.

Together, our ELL findings further inform the nuances of eye-cursor coordination in digital interactions: 1) like real-world Pick-ups, the eyes wait until the Pick-up happens then quickly leave, and 2) unlike real-world Drop-offs, the eyes dwell at the Drop-off site well beyond the end of the Transport.

2.5 Discussion

In this preliminary investigation, we show that eye-cursor coordination during a specific form of digital object interaction obeys constraints similar to eye-hand coordination during real-world interactions. Specifically, we find the eye dwells on or near the site of a digital interaction for at least 500 ms, almost identical to the amount of time others report in real-world interactions with physical objects. This initial finding demonstrates the potential utility of webcam eye-tracking collected from online, remote, crowdsourced participants as a tool for capturing rich, meaningful, and ecologically-valid visuomotor data.

Our study was designed to make the comparison of digital to real-world interactions as valid as possible. Thus, we adapted a previously reported real-world task (Lavoie et al., 2018) where the 500 ms minimum dwell time per interaction had previously been quantified. In our digital adaptation of this real-world cup-transfer task, we asked crowdsourced, online participants to perform 50 trials of an 8-movement drag-and-drop sequence (see Figure 2.1). While performing this digital interaction task, participants' cursor and gaze positions on the screen were monitored via the Labvanced experiment platform (Finger et al., 2017) using their own computer webcams. As described throughout this study, there are significant challenges to collecting webcam eye-tracking data, and as such, a major objective for this project was to assess its feasibility as a tool for quantifying patterns of dynamic eye-cursor coordination.

Tentatively, and with the caveat that substantial preprocessing was required, we believe that the eye data quality in this task was sufficient to explore eye-cursor coordination for this specific digital context. First, as reported by other research groups collecting online data (i.e. Semmelmann and Weigelt, 2018; Semmelmann et al., 2017; Yang and Krajbich, 2021), we experienced high rates of data exclusion (>40% of participants were not included in analysis, predominantly due to eye-data quality issues) even though we imposed restrictions on the hardware and internet connection of eligible participants. Second, for the participants who were included in the analysis, the eye data required extensive processing. This included reducing the spatial dimensionality from the (x,y) coordinate frame of the screen to 4 data-driven AOI bins (see Figure 2.2 for a representative subject), which were then mapped to the 4 task-relevant interaction locations. Given this approach departs from conventional eye-tracking analysis, we confirmed its sensitivity by testing if the distribution of gaze to each of these task-relevant AOIs followed the predictions imposed by task demand. Specifically, we show that participants fixated more on interaction locations that were relevant to the current movement (i.e. the target where you were clicking an object and the target where you were dragging it to) than to locations that were not relevant to that movement (see Figure 2.3, and Appendix B - Supplementary Materials for the complementary spatial investigation).

It is important to acknowledge that our successful collection and preliminary validation of webcam eye-tracking is in and of itself a significant contribution. With the uncontrolled nature of online tasks, and a technology that relies on consumer-grade hardware, there are many opportunities for noise or error to prevent successful data collection. We worked hard to minimize dropout due to hardware and internet issues by imposing very clear requirements, stated during crowdsourcing and checked during the study initialization. Then, we spent considerable time developing clear and transparent instructions to assist with participant retention. We provided detailed information about potential privacy concerns as well as video and interactive walk-through demonstrations of the task to promote participant understanding (see Appendix B - Supplementary Materials). As described above, our data was then processed using a k-means clustering technique. While clustering the eye-data was a successful approach for this study, it succeeded in part because our task-relevant AOIs were static and spatially distributed. Importantly, this means this approach will not be as successful or even possible for tasks with dynamic AOIs or AOIs that are close together. We further discuss these and other important limitations in Section 2.6 - Limitations and Future Directions, below.

Our investigation of gaze distribution (see Figure 2.3 for a temporal assessment, and Appendix B - Supplementary Materials for a spatial assessment) gave us sufficient confidence in the quality of the eye-tracking data to pursue our primary question of whether or not digital eye-cursor coordination in this task would follow similar patterns to real-world eye-hand coordination. Examining visuomotor coordination in tasks designed to promote natural, self-paced behaviours requires the task first be broken into its constituent interactions and then that those interactions be broken into the discrete phases of interaction. Adopting the segmentation strategy employed by Lavoie et al. (2018) we identified an important distinction between real and digital object interactions: despite both forms of movement being self-paced, digital objects are transported in about 200 ms (see Figure 2.4), 4 to 5 times faster than real objects are moved in the real world. Given these movements, it was impossible that the exact pattern of eye-hand coordination observed in the real world would be preserved during digital interactions. That is, during real-world interactions, when transporting an object between locations, the eye will "look ahead" to the drop-off site about 500 ms before the hand and object arrive. But, as just explained, during digital interactions, the *entire* transport lasts about 200 ms, meaning the eye cannot look ahead in the same way.

Remarkably, our results suggest that the visuomotor system preserves the overall interaction eye-dwell time of at least 500 ms by flexibly adapting the pattern of fixations. Specifically, we quantified eye-cursor latencies around both the Pick-up (cursor clicked to start dragging) and Drop-off (release of cursor click to stop dragging) events. At each event, we calculated how long the eye was looking at the location prior to the event (eye arrival latency, or EAL) and how long the eye remained looking at the location after the event (eye leaving latency, or ELL). As depicted in Figure 2.4, for a digital Pick-up, the pattern of gaze is almost identical to a real-world interaction: the eyes arrive at the location 400-500 ms before and stay for about 100 ms after. Given the speed of the Transport phase, eye latencies during digital Drop-off are significantly different from the real world. The eyes arrive around 100 ms before the event, but surprisingly, linger for 400-500 ms after the digital object has been released. With consideration to the preliminary nature of our investigation, we take this as the most important finding in our study: gaze allocation during a digital, self-paced, drag-and-drop object interaction flexibly adapts to the drastically different mechanics of movement to ensure about 500 ms of visual information is acquired from the beginning and end of each interaction movement. Thus, this perfectly aligns with the take-home message of real-world interactions, but the route by which it is achieved is significantly different.

Given this initial evidence for the persistence of ∼500 ms of eye fixation towards objects we interact with across real and digital domains (for at least this specific context), what insights might this offer for UX design principles? First, it suggests that as particular digital experiences are being designed, there may be a fundamental lower limit on the pacing with which interactions can occur while respecting the natural cadence of visuomotor coordination. For example, a drag-and-drop movement will take at least one second if executed at a natural speed. A designer wanting to push these movements to be faster might consider rearranging the target locations to make them spatially contiguous, possibly allowing the Pick-up and Drop-off information to be gathered by a single fixation. On the flip side, a designer requiring particularly precise cursor interactions might consider spatially separating the targets, letting the extra distance provide additional time for targeting fixations to occur within the natural rhythms of the task. In both of these cases, these preliminary findings support the notion that a design achieves maximum efficiency not by being the fastest, but rather, by aligning with the demands of a visuomotor system that evolved to optimally coordinate movements (Cisek, 2022) on its own time scale.

A second design principle which is more indirectly revealed via this introductory study is the role that feedback plays in facilitating successful interactions. Unlike in the real world where a person using their body to interact with an object typically receives haptic feedback about the interaction, digital interactions rely almost exclusively on visual feedback for confirmation that an interaction is proceeding successfully. In the current study we attempted to boost the visual cues associated with interaction by changing the visual properties of targets based on the cursor position. But, we believe there is more work to be done exploring how additional modalities could move digital interactions toward their real-world counterparts. For example, adding sound cues relevant to interaction, or even more sensitive and dynamic visual cues on objects that are successfully being interacted with may liberate the eyes to move in an even more natural fashion. Some work in high leverage interactions like laparoscopic surgery (Panait et al., 2009) has shown the utility of this approach. As real and digital worlds become less siloed and share elements across spaces (e.g. virtual and augmented reality), we predict that these mixed interactions will also obey the ∼500 ms of viewing time and so too will benefit from the exploration of multimodal feedback cues.

2.6 Limitations and Future Directions

Our goal was to measure more ecologically-valid user experiences from a diverse population of participants using their own digital devices in familiar environments. Thus, we collected eye-tracking data from webcams on remotely recruited participants. This meant we sacrificed some experimental control and introduced more eye-tracking noise, leading to a number of notable limitations. Here we describe some of these limitations and for each, offer a future direction for how to test and improve the study.

First, we provide no formal validation of the accuracy of the webcam eye-tracking system. While we attempted to quantify the functional accuracy in both time (see Figure 2.3) and space (see Appendix B - Supplementary Materials) a future approach would be to conduct the same experiment under controlled lab conditions while simultaneously recording both webcam and lab-grade, high-resolution eye-tracking data. Of course, a shift to the lab would also remove some of the environmental confounds of remote participation (lighting, hardware differences etc.) and would therefore provide a best-case measure of the magnitude of accuracy difference between webcam and lab-grade systems. Thus, given the known accuracy reduction in the current study, we urge the reader to take these results as preliminary and interpret them with due caution, leaving formal validation as a future opportunity.

Second, the reduced accuracy of the webcam eye-tracking forced us to define and use four, large, mutually-exclusive AOI bins with a spatial distribution roughly matching the actual targets but susceptible to spatial skewing (see Figure 2.2 and Appendix B - Supplementary Materials). As a result, we are unable to identify exactly where within a defined AOI the gaze was focused. Our analysis assumes that a look within a particular AOI is actually a look toward the relevant target object - an assumption that follows from real-world tasks where gaze is anchored only to task-relevant objects (Hayhoe, 2000; Land & Hayhoe, 2001). In the Supplementary Materials (see Appendix B) we present our attempt to quantify some aspects of this assumption by examining the spatial dispersion of the eye data when transformed into the task space. As reported, this analysis suggests that our AOI-binning was useful (unlikely to result in eye-data being mislabelled) but also highlights the remaining noise (most eye data still falls outside of a task-defined AOI). Therefore, the assumption of the specific timing and location of eye-positions relative to AOI boundaries should also be tested in future work. That is, further study is needed to confirm that this pattern of gaze anchoring extends to digital, screen-based contexts, and is also true in uncontrolled, remote settings. Again, a future study is needed to directly test this assumption by conducting an in-lab experiment comparing webcam to lab-grade eye-tracking.

Third, and perhaps most importantly, the large-AOI approach detailed above presents a significant limitation in our ability to draw definitive conclusions about how digital interactions relate to real-world interactions. This means the results of our second research question should be treated with particular caution as they are based on unvalidated assumptions. As an example, we are unable to provide any quantifiable proof that the arrival and dwell time of eye data binned into four large screen-based AOIs is equivalent, or even a good proxy for, the arrival and dwell time of a real-world fixation to a real-world object. As mentioned above, the spatial dispersion analysis of eye data clusters presented in the Supplementary Materials (see Appendix B) provides some quantified context for the general validity of this approach, but, again, a potential future solution would be to conduct a laboratory eye-tracking experiment benchmarked against a gold-standard, high-resolution eye tracker. Moreover, this future study could also explore a range of real-world and digital tasks to make the connections between them more clear. As an example, a more consistent real-world variant of the screen-based drag and drop task we employed here would have participants move and slide an object across a table, rather than lift and place as was done in the previous real-world experiment we used as inspiration (Lavoie et al., 2018).

Finally, the previous point about task variety highlights a limitation of our approach regarding its wider generalizability. To be clear, we only examined a single digital-interaction task that was designed to mimic only a single real-world object interaction task, and as such, our claims of this exposing a general property of the required fixation time for successful interaction (∼500 ms) should be tempered. As with how the body can interact with objects in the real world, during digital interactions there are a multitude of ways people can interact with digital objects. Here we limited ourselves to only one form of interaction (drag-and-drop) and as such, any conclusions we draw may only apply to that particular case. It is possible, or even likely, that other styles of interaction (e.g. point-and-click, swipe, hover) would result in different patterns of gaze behavior.

Taken together, these limitations highlight the need to interpret these results as exploratory and further validate some of the key assumptions. That being said, while accounting for these limitations, this study does suggest the promising power of remote data collection for ecologically-valid, user-centered, visuomotor research. This study also offers practical and automated methods (albeit task-specific) for extracting gaze patterns from unreliable data. Here, those methods provide sufficient quality to conduct an exploratory investigation of eye-cursor coordination under these more challenging conditions. In doing so, we believe we provide encouraging, though preliminary, evidence for some basic principles of gaze behavior during digital interactions. We think these results are exciting to drive future research focused on both validating the webcam results with in-lab experimentation, and exploring visuomotor coordination across the broader digital interaction space.

2.7 Conclusion

We believe our study introduces a potential new approach to user testing that accounts for aspects of UX that are rarely considered. Specifically, by adopting fully-remote testing of participants from the comfort of their own home, using their own hardware, we are actually testing users in the environment and context in which they'd typically encounter digital products. Moreover, our unique approach of harvesting gaze and movement data and automatically converting the data into objective metrics of experience gives rise to previously untapped insights. Here, in a preliminary investigation, aided by advances in webcam eye-tracking (Finger et al., 2017), we used this new level of insight to compare coordination across real and digital worlds, but this information may be valuable in innumerable contexts. The repertoire of digital interactions extends well beyond clicks and drags to points, swipes, flicks, pinches, taps and any number of other actions. Each of these is likely to be accompanied by a stereotyped pattern of natural visuomotor coordination which, when studied through the lens of gaze and movement behaviour, can help refine design processes. Already the benefits of this approach are being seen in real-world applications where scientists are better able to assess the movements of prosthetic limb users (Hebert et al., 2019) with the goal of helping those patients achieve more functionality in their activities of daily living.

It turns out, people *don't* move in mysterious ways. Instead, there are particular strategies for effective interactions that are true across drastically different environments. Like Fitts found for speed accuracy tradeoffs (Fitts, 1954), which retain their relationship on land (Fitts, 1954), underwater (Kerr, 1973), and in space (D. Newman & Lathan, 1999), here we report that across real and digital interactions, properties of eye-hand and eye-cursor coordination remain constant. By adapting design principles to align with these invariant properties of human performance we stand to improve the user's experience.

2.8 References

Arora, K. (2019). Eye-head-hand coordination during visually guided reaches in head-unrestrained macaques. *Journal of Neurophysiology*, *122*, 1946–1961.

Bánki, A., de Eccher, M., Falschlehner, L., Hoehl, S., & Markova, G. (2022). Comparing online webcam- and laboratory-based eye-tracking for the assessment of infants' audio-visual synchrony perception. *Frontiers in Psychology*, *12*, 733933.

Bieg, H.-J., Chuang, L. L., Fleming, R. W., Reiterer, H., & Bülthoff, H. H. (2010). Eye and pointer coordination in search and selection tasks. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10*, 89.

Bowman, M. C., Johannson, R. S., & Flanagan, J. R. (2009). Eye–hand coordination in a sequential target contact task. *Experimental Brain Research*, *195*(2), 273–283.

Butsch, R. L. C. (1932). Eye movements and the eye-hand span in typewriting [Place: US Publisher: Warwick & York]. *Journal of Educational Psychology*, *23*(2), 104–121.

Cisek, P. (2022). Evolution of behavioural control from chordates to primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1844), 20200522.

Danion, F. R., & Flanagan, J. R. (2018). Different gaze strategies during eye versus hand tracking of a moving target. *Scientific Reports*, *8*(1), 10059.

Deconinck, F. J. A., van Polanen, V., Savelsbergh, G. J. P., & Bennett, S. J. (2011). The relative timing between eye and hand in rapid sequential pointing is affected by time pressure, but not by advance knowledge. *Experimental Brain Research*, *213*(1), 99–109.

Deng, S., Chang, J., Kirkby, J. A., & Zhang, J. J. (2016). Gaze–mouse coordinated movements and dependency with coordination demands in tracing. *Behaviour & Information Technology*, *35*(8), 665–679.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: A unified JavaScript framework for online studies. *International Conference on Computational Social Science*, *1*(1), 1–3.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 11.

Fraser, A., Gattas, S., Hurman, K., Robinson, M., Duta, M., & Scerif, G. (2021, June 22). Automated gaze direction scoring from videos collected online through conventional webcam.

Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1436), 2435–2440. Retrieved September 13, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1690464/

Gagné, N., & Franzen, L. (2023). How to run behavioural experiments online: Best practice suggestions for cognitive psychology and neuroscience. *Swiss Psychology Open: the official journal of the Swiss Psychological Society*, *3*(1), 1.

Grossman, T., & Balakrishnan, R. (2004). Pointing at trivariate targets in 3d environments. *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, 447–454.

Guo, Q., & Agichtein, E. (2010a). Ready to buy or just browsing?: Detecting web searcher goals from interaction data. *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10*, 1–10.

Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, *7*(1), 43–64.

Hebert, J. S., Boser, Q. A., Valevicius, A. M., Tanikawa, H., Lavoie, E. B., Vette, A. H., Pilarski, P. M., & Chapman, C. S. (2019). Quantitative eye gaze and movement differences in visuomotor adaptations to varying task demands among upper-extremity prosthesis users. *JAMA Network Open*, *2*(9), e1911197.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. v. d. (2011, September 22). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Huang, J., White, R., & Buscher, G. (2012). User see, user point: Gaze and cursor alignment in web search. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1341–1350.

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye–hand coordination in object manipulation. *The Journal of Neuroscience*, *21*(17), 6917–6932.

Karousos, N., Katsanos, C., Tselios, N., & Xenos, M. (2013). Effortless tool-based evaluation of web form filling tasks using keystroke level model and fitts law. *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, 1851.

Kerr, R. (1973). Movement time in an underwater environment. *Journal of Motor Behavior*, *5*(3), 175–178.

Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, *25*(3), 296–324.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities. *Vision Research*, *41*(25), 3559–3565.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328.

Lavoie, E. B., Valevicius, A. M., Boser, Q. A., Kovic, O., Vette, A. H., Pilarski, P. M., Hebert, J. S., & Chapman, C. S. (2018). Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. *Journal of Vision*, *18*(6), 1–20.

Liebling, D., & Dumais, S. (2014). Gaze and mouse coordination in everyday work. *UBICOMP ADJUNCT '14*, 10.

Lin, C. J., & Ho, S.-H. (2020). Prediction of the use of mobile device interfaces in the progressive aging process with the model of fitts' law. *Journal of Biomedical Informatics*, *107*, 103457.

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human–Computer Interaction*, *7*(1), 91–139.

MacKenzie, I. S., & Buxton, W. (1992). Extending fitts' law to two-dimensional tasks. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92*, 219–226.

Mackenzie, I. S., Zhang, S. X., & Soukoreff, R. W. (1999). Text entry using soft keyboards. *Behaviour & Information Technology*, *18*(4), 235–244.

McGuffin, M. J., & Balakrishnan, R. (2005). Fitts' law and expanding targets: Experimental studies and designs for user interfaces. *ACM Transactions on Computer-Human Interaction*, *12*(4), 35.

Navalpakkam, V., Jentzsch, L., Sayres, R., Ravi, S., Ahmed, A., & Smola, A. (2013). Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. *Proceedings of the 22nd international conference on World Wide Web - WWW '13*, 953–964.

Newman, D., & Lathan, C. (1999). Memory processes and motor control in extreme environments. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *29*(3), 387–394.

Ngo, V., Gorman, J. C., De la Fuente, M. F., Souto, A., Schiel, N., & Miller, C. T. (2022). Active vision during prey capture in wild marmoset monkeys. *Current Biology*, *32*(15), 3423–3428.e3.

Panait, L., Akkary, E., Bell, R. L., Roberts, K. E., Dudrick, S. J., & Duffy, A. J. (2009). The role of haptic feedback in laparoscopic simulation training. *Journal of Surgical Research*, *156*(2), 312–316.

Patla, A., & Vickers, J. (2003). How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental Brain Research*, *148*(1), 133–138.

Pelz, J., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, *139*(3), 266–277.

Rand, M. K., & Stelmach, G. E. (2010). Effects of hand termination and accuracy constraint on eye–hand coordination during sequential two-segment movements. *Experimental Brain Research*, *207*(3), 197–211.

Rodden, K., & Fu, X. (2007). Exploring how mouse movements relate to eye movements on web search results pages. *Workshop on Web Information Seeking and Interaction at SIGIR '07*, 29–32. Retrieved September 12, 2022, from https://storage.googleapis.com/pub-tools-public-publication-data/pdf/32735.pdf#page=33

Roy, N., Câmara, A., Maxwell, D., & Hauff, C. (2021). Incorporating widget positioning in interaction models of search behaviour. *roceedings of the 2021 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR '21)*, 10.

Sailer, U., Eggert, T., Ditterich, J., & Straube, A. (2000). Spatial and temporal aspects of eye-hand coordination across different tasks. *Experimental Brain Research*, *134*, 163–173.

Semmelmann, K., Hönekopp, A., & Weigelt, S. (2017). Looking tasks online: Utilizing webcams to collect video data from home. *Frontiers in Psychology*, *8*(1582), 1–11.

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451–465.

Seow, S. C. (2005). Information theoretic models of HCI: A comparison of the hick-hyman law and fitts' law. *Human–Computer Interaction*, *20*(3), 315–352.

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2011). Adaptive allocation of vision under competing task demands. *Journal of Neuroscience*, *31*(3), 928–943.

Smith, B. A., Ho, J., Ark, W., & Zhai, S. (2000). Hand eye coordination patterns in target selection. *Eye Tracking Research & Applications Symposium*, 117–122.

Valevicius, A. M., Boser, Q. A., Lavoie, E. B., Murgatroyd, G. S., Pilarski, P. M., Chapman, C. S., Vette, A. H., & Hebert, J. S. (2018). Characterization of normative hand movements during two functional upper limb tasks. *PLOS ONE*, *13*(6), e0199549.

van Donkelaar, P., & Staub, J. (2000). Eye-hand coordination to visual versus remembered targets. *Experimental Brain Research*, *133*(3), 414–418.

William Soukoreff, R., & Scott Mackenzie, I. (1995). Theoretical upper and lower bounds on typing speed using a stylus and a soft keyboard. *Behaviour & Information Technology*, *14*(6), 370–379.

Williams, H. E., Boser, Q. A., Pilarski, P. M., Chapman, C. S., Vette, A. H., & Hebert, J. S. (2019). Hand function kinematics when using a simulated myoelectric prosthesis [ISSN: 1945-7901]. *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*, 169–174.

Wisiecka, K., Krejtz, K., Krejtz, I., Sromek, D., Cellary, A., Lewandowska, B., & Duchowski, A. (2022). Comparison of webcam and remote eye tracking. *2022 Symposium on Eye Tracking Research and Applications*, 1–7.

Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, *16*(6), 1485–1505. Retrieved August 24, 2022, from https://journal.sjdm.org/21/210525/jdm210525.html

Chapter 3

Assessing webcam eye-tracking utility in binary choice decision-making

Abstract

As decisions require the gathering of relevant information, eye-tracking measures that capture the way visual information is typically acquired offer powerful indices of the dynamic decision-making process. This study is the second of a pair of studies that explore continuous measures of decision-making using remote, online tools in naturalistic settings. While cursor-tracking, used in the companion paper (Ouellette Zuk et al., 2023), enabled access to dynamic decision processes expressed during movement, in the present study, we now employ webcam eye-tracking to examine the dynamics of information gathering during decision making prior to movement initiation. Using three previously published binary choice tasks, we explored indices of decision difficulty in the gaze dynamics that would complement the motor measures in our companion paper. We find that harder choices elicit more eye dwells and longer final dwells, reflecting a decision resolution process that Ouellette Zuk et al. index during the final choice movement. Beyond this, we identify distinct gaze patterns

A version of this work has been published on a pre-print server (not peer-reviewed) as Bertrand, J. K., Zuk, A. A. O., & Chapman, C. S. (2023). Continuous Measures of Decision-Difficulty Captured Remotely: II. Webcam eye-tracking reveals early decision processing. *bioRxiv*, 2023-06. doi:10.1101/2023.06.06.543799.

uniquely employed in each task, revealing the utility and sensitivity of gaze metrics in illuminating the early difficulty-independent information gathering processes at play. Together, this paper series demonstrates the power of remote, online methods as tools for deeply understanding the complete, dynamic and continuous decision process, from the first glance to the final response.

3.1 Introduction

In the past, research on decision-making has primarily focused on the outcomes of decisions and has used discrete measures such as choice-outcomes, accuracy, or reaction time to understand the underlying cognitive processes involved (see Schuetz et al., 2019 for review). However, these approaches really only capture *what* decision was made with little or no information about *how* a particular decision was arrived at. To address this gap, recent research has turned to the motor system as a means to investigate the dynamics of decision-making. Actions that move through space have been shown to track decision processes that unfold across time (Cisek & Kalaska, 2010; Dotan et al., 2019; Gallivan et al., 2018; Wispinski et al., 2020).

One of the most accessible ways to record movement data is through cursor movements on a screen. For example, 2-D computer-mouse movements provide a sensitive, flexible, and scalable method to study decision dynamics (Faulkenberry et al., 2016; Freeman, 2018; Hehman et al., 2015; Koop and Johnson, 2013; Moher and Song, 2014; Stillman et al., 2018, and many more). Moreover, the widespread adoption of cursor-based technologies (computers, tablets, smartphones) and the recent availability of online experiment generation tools (e.g., Labvanced, Finger et al., 2017; Gorilla, Anwyl-Irvine et al., 2020; lab.js, Henninger et al., 2022; jsPsych, Leeuw et al., 2023; PsychoPy, Peirce et al., 2019) means that now, more than ever, we can collect information about decision dynamics from larger, more diverse samples and from people in more ecologically-valid contexts (i.e., remote, online data collection). The current study is the second half of a two-article set aimed at mapping the possibilities and limitations of using remote data collection during cursor-based decision tasks. In the first study (Ouellette Zuk et al., 2023), we focus on testing the robustness of remote data collection for understanding dynamic decision-making, showing that nuanced details of decision processes are available not only in computer-mouse movements but are also clearly evident and sometimes even stronger when the same task is deployed on tablets and smartphones.

But, our first study is literally only half the story. While cursor-tracking reveals valuable insights into the decision process once movement begins, it cannot track details of the perceptual processes that occur prior to a movement toward a particular choice being initiated. While reaction time reliably tracks decision difficulty in this early phase (e.g., Palmer et al., 2005; Schouten and Bekker, 1967) it fails to capture any of the constituent dynamics of how the decision process is evolving. In these crucial early moments of a decision, a person is gathering the necessary information from their environment to inform their choice. For tasks involving visual stimuli in particular, this information gathering is primarily mediated by eye movements. Therefore, eye-tracking offers a unique opportunity to investigate the decision process at an earlier stage, revealing how individuals extract information and make decisions based on where they focus their gaze.

Given this, it is not surprising that eye-tracking has long been a prominent method in decision-making research. Extensive work has shown the interconnectedness of gaze and choice, where gaze patterns can both reflect and bias choices (Glaholt & Reingold, 2009; Glaholt et al., 2009; Shimojo et al., 2003). Moreover, eye movements have been shown to actively sample the world in a way that adaptively maximizes the informative value of fixations (Cassey et al., 2013; Gottlieb, 2018). Analyzing these fixation sequences that precede a decision has provided valuable insights, revealing that the location, duration, and pattern of fixations serve as indices of the relative competition between choice options (Krajbich & Rangel, 2011; Krajbich et al., 2010). Findings like these have challenged classic decision-making theories (e.g., evidence accumulation; Gold and Shadlen, 2007; Ratcliff and Rouder, 1998; P. L. Smith and Vickers, 1988) to reconcile the important role the eyes play in information sampling. This has led to the development of gaze-aware decision models such as the attentional drift diffusion model (aDDM; Krajbich et al., 2010), Decision Field Theory (Busemeyer & Townsend, 1993), and the gaze cascade model (Shimojo et al., 2003). These models assume an option receives more evidence when gazed at, acting like an amplifier for the attended option (Krajbich, 2019). Moreover, the aDDM has been extended beyond preferential choice contexts, encompassing different stimuli types (e.g., numeric information vs pictorial images; Krajbich et al., 2012) and choice domains (e.g., risky and social choices; S. M. Smith and Krajbich, 2018). Collectively, these theoretical advancements and empirical contributions provide a foundation for comprehending how the eyes sample spatially-distributed information, emphasizing the value of eye-tracking as a method to better understand how decisions unfold. In the context of the current study, they show that eye-tracking is almost a perfect complement to cursor-tracking in its ability to fill in the gap of decision dynamics during the earliest stages of a choice being made.

A major drawback shared by all of the aforementioned eye-tracking studies is their confinement to laboratory settings. Recently, however, the use of webcam eye-tracking has emerged as a promising avenue in bridging the gap between controlled laboratory experiments and data collected in a wide range of environments (e.g., Bertrand and Chapman, 2023; Stone and Chapman, 2023; Yang and Krajbich, 2021). Admittedly, webcam eye-tracking is still a method in its infancy and has notable limitations in both temporal and spatial accuracy (Bánki et al., 2022; Semmelmann & Weigelt, 2018; Yang & Krajbich, 2021). Despite these challenges, it offers a distinct advantage by capturing gaze patterns in ecologically-valid settings such as within the participants' own homes and on their personal devices. Furthermore, it can be argued that - with the overall trend toward increasing digital and screen based interactions - computerized decision-making tasks like the one used by Krajbich et al. (2010), now closely resemble everyday, real-world decisions. Thus, by shifting these tasks to a more naturalistic context, such as the participants' homes, a more comprehensive exploration of authentic, real-world visual behaviours and decision-making processes becomes possible. Additionally, the remote nature of this method allows for scalable data collection beyond the limitations of a laboratory, while also providing access to more diverse, global populations (Aguinis et al., 2021; Johnson et al., 2021).

Therefore, in this study we employ webcam eye-tracking - in part to explore its potential and limits as a method - and theoretically to explore the relationship between decision difficulty and gaze behaviour patterns. As previously mentioned, the current study is a companion paper to Ouellette Zuk et al. (2023), with both studies intentionally sharing the same experimental design. This design aimed to replicate and extend three unique and previously published mouse-tracking based decision-making tasks. These tasks were deliberately chosen to cover a range of decision domains including objective perceptual judgments (Numeric Size-Congruity, Faulkenberry et al., 2016), semi-subjective conceptual judgements (Sentence Verification, Dale and Duran, 2011) and subjective preference judgements (Photo Preference, Koop and Johnson, 2013). The tasks also varied in terms of stimulus characteristics, encompassing numerical digits, written statements, and photos. By employing webcam eye-tracking, we not only gain insights into how the decision context varies across tasks but also how differences in the presentation and distribution of information across space affect the decision-making process, something that cannot be solely obtained through mouse-tracking. Thus, replicating Ouellette Zuk's design with eye-tracking will not only allow us to explore the rich, dynamic decision process earlier in time, beginning before movement initiation, but also explore how this process presents across decision contexts and with different distributions of decision information in the display. In doing so, we also demonstrate the utility of remote data collection in general and webcam eye-tracking in specific as a tool for capturing this rich readout of decision-making from participants in their own environments using their own devices.

3.2 Results

This study replicates and extends our cursor-tracking-focused companion paper (Ouellette Zuk et al., 2023). Both studies employed three binary choice tasks where participants indicated their choice through cursor movements: a Sentence Verification task (Dale & Duran, 2011), a Numeric-Size Congruity task (Faulkenberry et al., 2016), and a Photo Preference task (Koop & Johnson, 2013). Each task was designed and analyzed to produce Easy and Hard trials (see Figure 3.1): For the Sentence Verification task, based on previous work (Dale & Duran, 2011), participants indicated if a simple statement was true or false. On Easy trials, the statements were true and not negated (e.g., 'Cars have tires') and on Hard trials the statements were true and negated (e.g., 'Cars do not have wings'). For the Numeric-Size Congruity task, based on previous work (Faulkenberry et al., 2016), participants indicated which of two digits had a higher numeric value. On Easy trials the size and numeric value were congruent (e.g., 2 vs. 8), and on Hard trials size and value were incongruent (e.g., 2 vs. 8). Finally, for the Photo-Preference task, based on previous work (Koop & Johnson, 2013), participants indicated which of two photos they preferred. On Easy trials, one photo had low pleasantness while the other had high pleasantness and on Hard trials both photos had high pleasantness. Our companion paper successfully replicated the main finding from the original task publications (i.e., Dale and Duran, 2011; Faulkenberry et al., 2016; Koop and Johnson, 2013) that responses on Hard trials take longer than Easy trials, while also showing how decision difficulty affects cursor movements. Here we predict that we will also replicate the finding that Hard trials generate longer response times than Easy trials and investigate how decision difficulty affects webcam-tracked gaze behaviour.

3.2.1 Response Time: Hard decisions take longer than easy decisions

We use Response Time as a single, broad measure to capture the duration from the presentation of choice options to the point of response (marked by the cursor entering the selected option). It encompasses both reaction time and movement time as described in our companion paper (Ouellette Zuk et al., 2023). Replicating our previous work, and confirming our key prediction, across all three tasks we observed a consistent Difficulty effect (see Figure 3.1). Specifically, Hard trials required significantly more time than Easy trials (Sentence Verification: $t(96) = 21.0$, $p < .001$, $d = 2.13$; Numeric-Size Congruity: $t(89) = 8.01$, $p < .001$, $d = 0.84$; Photo Preference: $t(96) = 7.95$, $p < .001$, $d = 0.81$).

3.2.2 Proportion of Trials: Unique task demands drive unique gaze behaviours

We began our examination of gaze patterns within each task by characterizing and analyzing the most frequently observed dwell patterns (see Figure 3.1). For our analysis, we defined a dwell as a continuous gaze on an area of interest, lasting at least 100 milliseconds within the expanded boundaries of that area (see Figure 3.3). A dwell ended if the gaze shifted outside that area for more than 100 milliseconds. To identify the most common gaze patterns we conducted separate RMANOVAs for each task. These analyses aimed to assess when and how often the eyes dwelled on particular AOIs. The following RMANOVAs include up to four factors to describe the dwell patterns observed: Dwell Count, Difficulty, First Dwell Side, and Last Dwell Option. The Dwell Count factor consisted of four levels (1, 2, 3, and 4 or more), indicating the number of unique dwells during a trial. The Difficulty factor distinguished between trials classified as either Hard or Easy. The First Dwell Side factor described whether the initial dwell on a trial fell upon the Left or the Right choice option. Finally, the Last Dwell Option factor indicated whether the final dwell

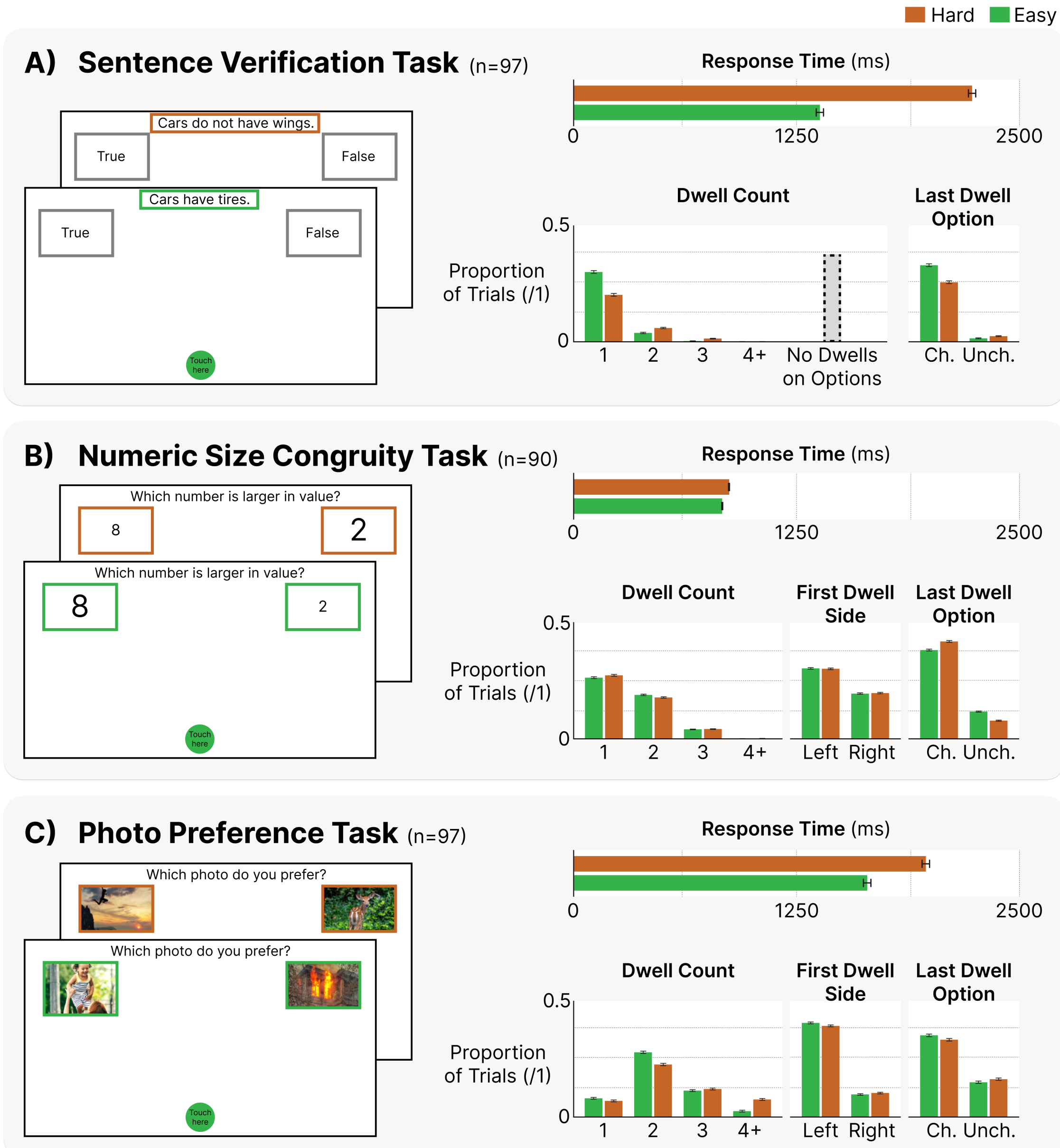


Figure 3.1: Examples of hard and easy decisions, alongside response time (horizontal bar graphs) and the proportion of trials (vertical bar graphs) results across the study's three tasks: A) Sentence Verification, B) Numeric Size Congruity and C) Photo Preference. Throughout, orange represents hard decisions, while green represents easy decisions. Error bars are the standard error of the difference between the hard and easy conditions. The proportion of trials data is presented as simplified marginal means, where each factor analyzed in the Repeated Measures ANOVA (RMANOVA) is presented independently across its levels (note (Ch.) means Chosen and (Unch.) means Unchosen) The Sentence Verification task includes an untested but present gaze behaviour of No Dwells on Options, indicating the proportion of trials where only the sentence received a dwell.

of the trial was on the Chosen or Unchosen option.

Sentence Verification Task

Sentence Verification was unique in that many times (∼35% of all trials) participants' eyes dwelled only on the sentence and never on either choice option. We represent the proportion of these "No Dwells on Options" in Figure 3.1A, but it is not possible to examine them in our statistical analysis of choice option gaze patterns. Thus, we acknowledge their presence as a predominant gaze behaviour in Sentence Verification and proceed with the rest of the analysis looking only at trials having at 1 or more dwells.

A 2 (Last Dwell Option) x 2 (Difficulty) x 4 (Dwell Count) RMANOVA of the proportion of trials within the Sentence Verification task revealed a significant three-way interaction between the tested factors (see Table 3.1; $F(1.27,121.97) = 53.6$, $p < .001$, $\eta_p^2 = 0.36$). This interaction was interrogated further by splitting the data into the much more common trials that ended with a dwell on the Chosen option (∼60%) and those that more rarely ended with a dwell on the Unchosen option (∼5%). For each of these groups we ran separate 2 (Difficulty) x 4 (Dwell Count) RMANOVAs.

When we look at the three-way interaction follow-up RMANOVA for Last Dwell on Chosen we reveal a significant interaction between Difficulty and Dwell Count ($F(1.29,123.74) = 56.4$, $p < .001$, $\eta_p^2 = 0.37$), as well as significant main effects (Dwell Count: $F(1.33,127.76) = 281.9$, $p < .001$, $\eta_p^2 = 0.75$; Difficulty: $F(1,96) = 41.1$, $p < .001$, $\eta_p^2 = 0.30$). We further followed up the two way interaction with four 1-factor RMANOVAs - one for each level of Dwell Count. Here, we see significant Difficulty effects in the proportion of trials at each level of Dwell Count. For last dwell on Chosen one-dwell trials, we see a higher proportion of Easy trials than Hard trials ($p < .001$, $M_{Hard-Easy}$ = -0.104), while the reverse pattern of a higher proportion of Hard than Easy trials is revealed when there are two, three, and four or more dwells (all p's $< .05$, and $M_{Hard-Easy}$ range from 0.00215 to 0.0192). In general, this suggests that for trials that show the commonly-occurring Last Dwell on Chosen pattern, Harder trials result in more dwells than Easy trials.

When the Last Dwell fell on the Unchosen option (which occurred rarely, $M_{Unchosen}$ = 0.0434), we observed differences in the proportion of trials driven by both Difficulty (F(1,96) = 5.19, p = .0249, η_p^2 = 0.051) and the Dwell Count (F(1.17,112.37) = 61.16, p < .001, η_p^2 = 0.39), but not their interaction (F(1.22,116.65) = 1.99, p = 0.1582, η_p^2 = 0.020). The proportion of Hard trials was slightly higher than that of Easy trials ($M_{Hard-Easy}$ = 0.0024), and all pairwise comparisons between the number of dwells yielded significant results (all p's < .05), with the proportion of trials following a pattern of one-dwell > two-dwells > three-dwells > four or more dwells. The results suggest that participants have very few trials with multiple dwells in this task and that the relatively rare gaze pattern where the eyes end on the Unchosen option occurs slightly more on Hard trials than Easy trials.

Overall, Figure 3.1A highlights the general gaze patterns elicited by the Sentence Verification task: while sometimes there was no gaze upon the choice options, when it did happen, it was usually only once, and almost always on the chosen option. However, in the relatively infrequent number of trials where the gaze dwelled more than one time on the options, this occurred more often on hard trials.

Numeric-Size Congruity Task

We explored the proportion of trials in Numeric-Size Congruity with a 2x2x2x4 RMANOVA (Last Dwell Option x First Dwell Side x Difficulty x Dwell Count). Two significant three-way interactions were revealed (these were the highest order significant interactions): Last Dwell Option x Difficulty x Dwell Count (see Table 3.1; F(1.85,164.29) = 22.2812, p < .001, η_p^2 = 0.20) and Last Dwell Option x First Dwell Side x Dwell Count (F(2.26,201.05) = 18.0261, p < .001, η_p^2 = 0.17).

To follow-up the Last Dwell Option x Difficulty x Dwell Count interaction, like the Sentence Verification follow-up, we looked at Difficulty x Dwell Count at each Last

Dwell Option level separately. Again, participants were much more likely to end the trial by looking at the Chosen ($\sim$80%) as compared to the Unchosen ($\sim$20%) option. Following up the three-way interaction for the more common last dwell on Chosen gaze pattern, we uncovered a two-way interaction between Difficulty and Dwell Count ($F(1.75{,}155.67) = 10.1, p < .001$, $\eta_p^2 = 0.10$). The further follow-ups at each level of Dwell Count found that the interaction was driven primarily by a significant Difficulty effect on one-dwell trials only. That is, there is a greater proportion of single, Chosen option dwells on Hard trials compared to Easy trials ($p < .001$, $M_{Hard-Easy(1Dwell)} =$ 0.0353, $\eta_p^2 = 0.18$). Overall, this result highlights the high proportion of trials where the last dwell ends on the chosen option, while revealing a subtle difficulty effect in single dwell trials. We believe this difficulty effect is best understood after also explaining the pattern of behaviour on Last Dwell on Unchosen trials.

For the less common Last Dwell on Unchosen trials, the 2x4 follow-up RMANOVA revealed a significant two-way interaction ($F(1.86{,}165.17) = 26.9$, $p < .001$, $\eta_p^2 = 0.23$). Further follow-ups highlight that this is an effect driven by a significantly greater proportion of Last Dwell on Unchosen trials in Easy one and two-dwell cases than Hard (both p's $< .001$, $M_{Hard-Easy(1Dwell)} =$ -0.0261, $\eta_p^2 = 0.36$; $M_{Hard-Easy(2Dwells)}$ $=$ -0.0133, $\eta_p^2 = 0.16$). Last Dwell on Unchosen is a less common gaze pattern as compared to Last Dwell on Chosen, but when present, it's more likely to happen on Easy trials where there's only one or two dwells to the choice options. So, why are we seeing fewer dwells on harder trials, opposite to what we saw in the other tasks? We speculate this has to do with a unique property of the Numeric-Size Congruity task where what makes a specific trial hard is the physical size of the target, which we feel is likely related to its visual discriminability. Specifically, on Hard trials, your eyes have landed on a physically small but numerically large number. This smaller character likely takes additional time to resolve. During this time, however, it is possible that you are also processing information from the other target location (E. Stewart et al., 2020). On these Hard trials, the other numeral at this peripheral location is numerically small and physically large. We think that it may be easier to resolve this peripheral, larger target, eliminating the need for a second fixation. If we take the mirror scenario, on an Easy trial your eyes land on a numerically large digit that is also physically large. Resolving this larger stimulus occurs quickly. But, the peripheral target in these cases is physically small. We speculate that on some trials there is too much uncertainty about the identify of the peripheral stimulus which in turn drives a second dwell to its location. The net result of this is that, on some Hard trials your linger at the first location and don't need a second fixation while on some Easy trials you can more quickly leave the first location but feel the need to take a look at the second location.

Returning back to the second three-way interaction (Last Dwell Option x First Dwell Side x Dwell Count) that emerged from the omnibus RMANOVA, we chose to first explore this interaction at each level of First Dwell Side separately. In general, whether looking at trials where the gaze first fell on the Left option (which occurred ∼60% of the time) or the Right option (which occurred ∼40% of the time), the results are relatively similar. In both cases, a significant two-way interaction emerges between Last Dwell Option and Dwell Count (First Dwell Left: $F(1.66,147.96) = 80.3$, $p < .001$, $\eta_p^2 = 0.47$; First Dwell Right: $F(1.74,154.92) = 149$, $p < .001$, $\eta_p^2 = 0.63$). Further follow-ups for each case (at each level of Dwell Count) show the same pattern: the proportions of Last Dwell on Chosen trials is significantly greater than the Last Dwell on Unchosen trials (all 8 final tests have p values $< .01$).

All together, Figure 3.1B summarizes these results with marginal means shown for simplicity. In the Numeric-Size Congruity task, most often there were only one or two gazes upon the choice options, and in both cases, the first gaze was most likely to start on the left while the last gaze was almost always on the chosen option. The difficulty effects in this task were more subtle. When the gaze only landed on the chosen option and stayed there, it was more likely that this happened on a hard trial than an easy trial. But, in the small proportion of trials where the last dwell was on the unchosen option, there were more one dwell and two dwell gazes on easy trials than hard trials. We speculate this difficulty effect has to do with peripheral processing and the visual discriminability of targets of different sizes.

Photo Preference Task

We tested the proportion of trials in the Photo Preference task with a 2x2x2x4 RMANOVA (Last Dwell Option x First Dwell Side x Difficulty x Dwell Count) and found a significant four-way interaction between Last Dwell Option, First Dwell Side, Difficulty and Dwell Count ($F(2.65,254.36) = 29.94$, $p < .001$, $\eta_p^2 = 0.24$). For Photo Preference trials the most dominant factor was the First Dwell side with first looks to the Left (∼80%) being much more common than first looks to the Right (∼20%). As such, our initial follow-ups to the omnibus RMANOVA involved performing a 3-factor RMANOVA at each of the two levels of First Dwell Side.

In the more commonly-occurring trials where the First Dwell started on the Left, the three-way follow-up test revealed a significant three-way interaction between Last Dwell Option, Difficulty and Dwell Count (see Table 3.1; $F(2.57,247.20) = 30.43$, $p < .001$, $\eta_p^2 = 0.24$). The additional follow up two factor RMANOVAs that were performed for each level of Last Dwell Option revealed further interaction effects between Difficulty and Dwell Count in both tests (Last Dwell Unchosen: $F(1.98,189.71) = 4.82$, $p = .009$, $\eta_p^2 = 0.048$; Last Dwell Chosen: $F(2.69,258.31) = 35.89$, $p < .001$, $\eta_p^2 = 0.27$). In the most commonly occurring First Dwell Left-Last Dwell Chosen case, the two-way interaction follow-ups for each level of Dwell Count revealed a difficulty effect only for two and four or more dwells. For First Dwell Left, Last Dwell Chosen, two-dwell trials, there was a greater proportion of Easy trials than Hard trials ($M_{Hard-Easy} = -0.0637$, $p < .001$, $\eta_p^2 = 0.44$), and an opposite pattern for four or more dwell trials ($M_{Hard-Easy} = 0.0312$, $p < .001$, $\eta_p^2 = 0.23$). In the less common First Dwell Left, Last Dwell Unchosen case, when further follow-ups were performed at each level of Dwell Count, we found the two-way interaction to be driven by significant differences in trials with one and four or more dwells. For one-dwell, Last Dwell Unchosen, First Dwell Left trials, there was a significantly greater proportion of Easy trials than Hard trials ($M_{Hard-Easy}$ = -0.00921, p = .00421, η_p^2 = 0.08) and the opposite pattern for four or more dwell trials ($M_{Hard-Easy}$ = 0.0138, $p < .001$, η_p^2 = 0.16). Taken together we see that Hard trials generally shift toward having more dwells (three or more) than Easy trials (one or two).

Shifting to the follow up analysis of the more infrequent trials where the First Dwell was to the Right, the only significant effects came from the main effects of Last Dwell Option and Dwell Count (Last Dwell Option: F(1,96) = 17.502, $p < .001$, η_p^2 = 0.15; Dwell Count: F(2.06, 197.43) = 24.566, $p < .001$, η_p^2 = 0.20). All the pairwise comparisons were significant, with a greater proportion of First Dwell Right and Last Dwell Chosen trials than First Dwell Right and Last Dwell Unchosen trials ($M_{Chosen-Unchosen}$ = 0.00791, $p < .001$), and, when the First Dwell started on the Right, a pattern of greater proportions of two-dwells than one-dwell than three-dwells than four or more dwells (all p's $\leq$ .01689).

Figure 3.1C highlights the dominant gaze patterns evident during the Photo Preference task. Most often, participants looked at each option at least once, almost always starting on the left and ending on whichever option they chose. Decision difficulty inflated the number of dwells, with participants more likely to make more dwells if the decision was harder.

3.2.3 Gaze Dynamics: Driven early by stereotyped information gathering, affected later by decision difficulty

We used our proportion of trials analyses to guide an in depth exploration of gaze dynamics for the most commonly occurring gaze patterns in each task. This ensured adequate statistical power and allowed us to fully represent the dramatic ways gaze patterns differed across tasks. The analyses presented in the current section focused on the temporal aspects of gaze patterns and specifically examined how these dynam-

Last Dwell (%, SD)	Dwells Count	Difficulty		Hard-Easy					
		Hard	Easy	*M*	*SE*	df	*F*	*p*	η^2_p
Sentence Verification									
Chosen (58.78%, 27.53%)	1	0.181	0.286	-0.104	0.013	1,96	64.8	***	0.40
	2	0.057	0.038	0.019	0.006	1,96	10.9	**	0.10
	3	0.015	0.006	0.009	0.003	1,96	11.5	**	0.11
	4+	0.002	5.52e-4	0.002	9.83e-4	1,96	4.8	*	0.048
Unchosen (0.043,0.0476)	-	0.006	0.004	0.002	0.001	1,96	5.2	*	0.051
Numeric-Size Congruity									
Chosen (80.10%,10.78%)	1	0.230	0.195	0.035	0.008	1,89	20.1	***	0.18
	2	0.144	0.143	0.002	0.006	1,89	0.08	n.s.	0.001
	3	0.042	0.042	4.32e-4	0.003	1,89	0.02	n.s.	0.000
	4+	0.003	0.003	4.82e-4	9.79e-4	1,89	0.24	n.s.	0.003
Unchosen (19.90%,10.78%)	1	0.044	0.070	-0.026	0.004	1,89	49.0	***	0.36
	2	0.034	0.048	-0.013	0.003	1,89	16.5	***	0.16
	3	0.002	0.001	0.001	8.09e-4	1,89	2.7	n.s.	0.029
	4+	0.00	0.00	0.00	0.00	1,89	NaN	NaN	NaN
Photo Preference (only Left Start; 79.66%, 22.43%)									
Chosen (55.17%, 19.54%)	1	0.032	0.029	0.003	0.005	1,96	0.44	n.s.	0.005
	2	0.087	0.150	-0.064	0.007	1,96	74.1	***	0.44
	3	0.091	0.088	0.003	0.007	1,96	0.16	n.s.	0.002
	4+	0.053	0.022	0.031	0.006	1,96	27.9	***	0.23
Unchosen (24.48%, 13.05%)	1	0.008	0.017	-0.009	0.003	1,96	8.6	**	0.082
	2	0.089	0.079	0.010	0.007	1,96	1.9	n.s.	0.019
	3	0.016	0.017	-0.001	0.004	1,96	0.13	n.s.	0.001
	4+	0.016	0.002	0.014	0.003	1,96	18.5	***	0.16

Table 3.1: Proportion of trials results from each task's three-way interaction involving Difficulty as a factor. Note. *p < .05; **p < .005; ***p < .0005

ics varied with decision difficulty. We operationalized the gaze dynamics into timing metrics to describe the onset and duration of a dwell, as well as the time from the end of the dwell to the end of the trial (where the trial ended upon the cursor entering the chosen option). In cases where the most frequent gaze pattern involved more than 1 dwell, we also measured the onset and duration of the second or last dwell, as well as the time between dwell events. Across the three tasks, we analyzed 21 gaze dynamics metrics. We conducted paired t-tests comparing the Hard and Easy Difficulty conditions for each metric. To account for the potential impact of multiple comparisons, we adjusted the significance level to $p = .05/21 = .00238$. We fully report the results of these twenty-one gaze metrics in Table 3.2.

Sentence Verification Task

The most common gaze behaviour observed during the Sentence Verification task was either no dwell on any choice option or a single dwell on the chosen option. As mentioned earlier, we did not investigate gaze dynamics for the no dwell on option trials. To examine the gaze behaviour on single dwell trials, we analyzed three metrics: First Dwell Onset, First Dwell Duration, and End Dwell to Response (see Figure 3.2A). In the Sentence Verification task, irrespective of decision difficulty, there was no difference in the duration of the dwell on the chosen option. However, for both First Dwell Onset, and the time from the end of the dwell up to the response (End Dwell to Response), Hard decisions elicited significantly longer latencies (see Table 3.2).

While we know that Response Times are longer for hard decisions (see above), these gaze dynamics suggest that the additional time is not simply due to an elongated timeline for all components of a Hard decision. Rather, in this particular task, where decision-relevant information resides outside of the choice options, more time is spent reading the sentence for Hard decisions before the gaze moves to the chosen options (approximately 700 ms). But, the chosen option itself does not require extensive

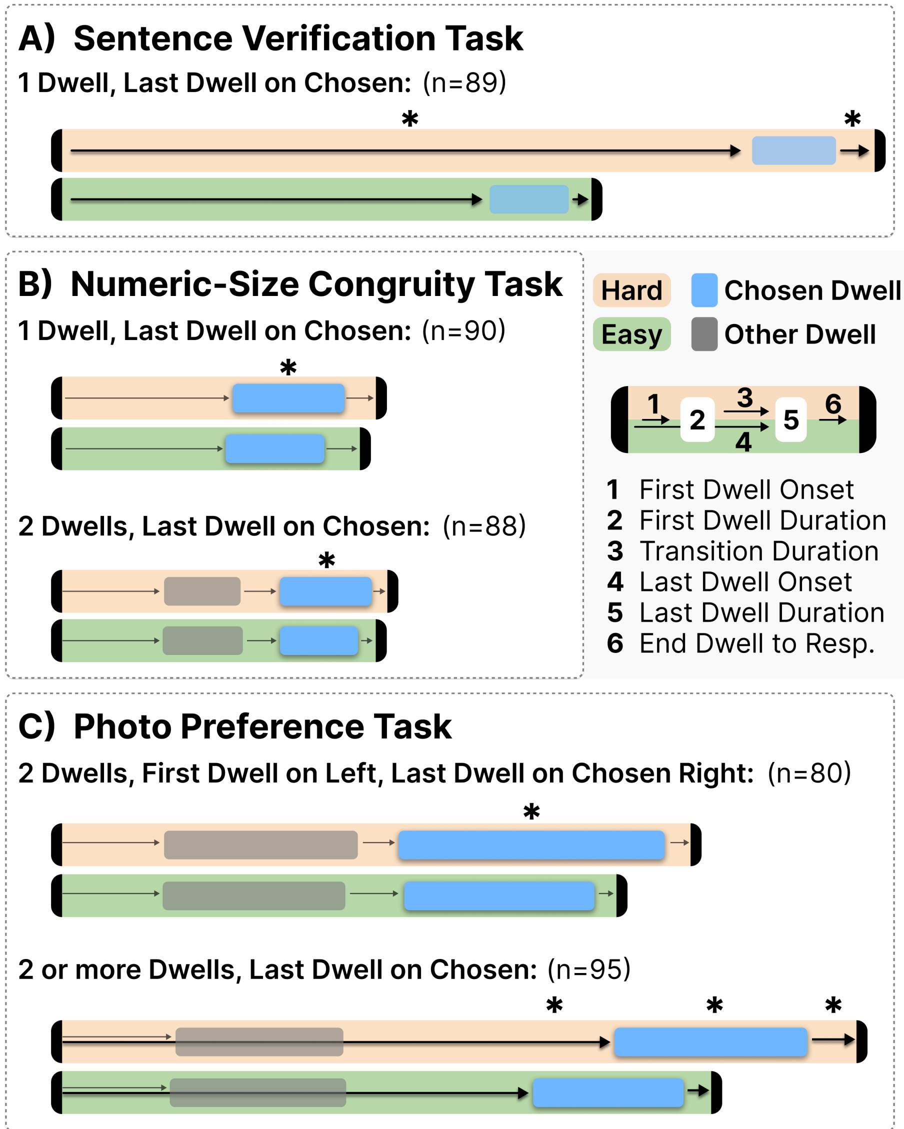


Figure 3.2: Gaze dynamics of the most common dwell patterns across the three decision tasks. Each pattern is depicted for hard and easy decisions, with every decision shown within a horizontal orange (hard) or green (easy) bar. All gaze patterns are aligned to the moment choice options are presented. Arrows are used to indicate the onsets and offsets of dwells, while the blue bar shows the dwell duration on the chosen option, and the grey bar shows the dwell duration on the other option for patterns where there is more than one dwell. Any significant differences between the gaze metrics of hard and easy trials are indicated with a star (*) and an opaque bar or thicker arrow. Insignificant differences are shown as transparent bars or thin arrows. For the Photo Preference 2 or more dwells plot, only the first and last dwell are depicted, but the results include trials where there would be additional dwells. Detailed means and statistics are available in Table 3.2.

viewing as the 'True' and 'False' options maintain a consistent position throughout the task. Hence this dwell time does not differ between Hard and Easy trials. However, after the gaze leaves the chosen option, a harder decision requires more time to complete compared to an easier decision. This suggests that there is further cognitive processing involved in resolving the decision before the mouse cursor lands in the chosen option. We discuss this finding in light of our parallel mouse-tracking study (Ouellette Zuk et al., 2023) in the Discussion.

Numeric-Size Congruity Task

For Numeric-Size Congruity we analyzed the two most prevalent gaze behaviours: a single dwell on the chosen option, and two dwells with the last dwell on the chosen option. Interestingly, in contrast to Sentence Verification, the effect of difficulty on gaze pattern timing in single dwell trials was entirely different. Figure 3.2B illustrates that the only significant difference between Hard and Easy trials was found in the dwell duration of the single fixation, with Hard trials eliciting longer dwells than Easy trials (see Table 3.2). In this task, since the decision-relevant information is located at the choice option, the effects of decision difficulty are primarily expressed in the time required to visually acquire this information.

Further insights about the decision-making process are gained when exploring the two-dwell pattern with the last dwell on the chosen option. In these trials, we segmented the pattern into 5 constituent metrics: time to first dwell, first dwell duration, time between dwells, second dwell duration and second dwell offset to response. Remarkably, only the duration of the second (and last) dwell showed a significant effect of decision difficulty (see Table 3.2). This indicates that in two-dwell trials, both Hard and Easy trials exhibit the same initial dwell gaze pattern up until the second, and chosen, choice option is viewed. Only at this point, does the gaze tend to dwell longer on average for Hard decisions, suggesting that additional time is needed to process and integrate the decision-related information obtained from both options.

Photo Preference Task

To capture the unique gaze behaviours specific to Photo Preference, we examined two distinct patterns: the highly stereotypic gaze pattern of two dwells with the first dwell on the left and the last dwell on the chosen (and right) option, as well as a broader set of patterns involving more than one dwell where the last dwell was on the chosen option. Similar to the Numeric-Size Congruity task, we analyzed five metrics for two-dwell, first-dwell-left, last-dwell-chosen trials. Once again, we found significant effects of difficulty only in the second (and last) dwell, with Hard trials resulting in dwells significantly longer than Easy trials (see Table 3.2). Figure 3.2C illustrates this pronounced difference, and we later discuss how the form of this decision information (colorful photos) and the type of decision required (preference) likely contribute to this finding.

Lastly, we also examined decision difficulty in Photo Preference across all gaze patterns involving more than one dwell (trials with two, three or four or more dwells) where the last dwell fell on the chosen option. Here, we still analyzed five metrics, but with adjustments: metrics were anchored to the 'last' dwell (whether it was the second, third, fourth or more), and the measure of time between the two dwells was replaced with the onset time of the last dwell relative to the start of the trial. This is depicted in Figure 2C. In these analyses, we observed significant differences between Hard and Easy trials in the duration of the last dwell, the onset of the last dwell, and the offset of the last dwell to the response (see Table 3.2). Hard trials consistently exhibited longer latencies in all three metrics. As these tests were performed on aggregated data from two, three, four, or more dwells, the onset of the last dwell aligns with the results from the proportion of trials, where Hard Photo Preference trials had a greater proportion of trials with more dwells compared to Easy trials. Once again, the duration of the last dwell provided evidence that it was not exclusively the 'second' dwell but rather the 'last' dwell affected by decision difficulty. Further, the impact of decision difficulty extended beyond the last gaze, as Hard decisions appeared to require additional time for resolution, consistent with the longer movement times in Ouellette Zuk et al.'s study (2023). Perhaps most interestingly, just like the Size-Congruity task, the timing of the first dwell was not impacted by decision difficulty. That is, in terms of both onset and duration, the first dwell is highly stereotyped, suggesting that decision competition doesn't really begin until all decision information has been sampled.

3.3 Discussion

In this study we collected data from a remote cohort of participants performing three different binary choice tasks while recording gaze behaviour via webcam eye-tracking. This paper serves as a companion to Ouellette Zuk et al. (2023) which details how decision-dynamics play out for the same three tasks when measured by mouse-tracking. Both papers rely on previously published works (Dale & Duran, 2011; Faulkenberry et al., 2016; Koop & Johnson, 2013) that identified trials with hard or easy decisions. Briefly the three tasks were: A Sentence Verification task where you determined whether a statement was true or false (difficulty was manipulated through sentence negation); A Numeric-Size Congruity task where you determined which of two digits was numerically larger (difficulty was manipulated through the congruence of physical and numerical size); and a Photo-Preference task where you determined which of two photos you preferred (difficulty was manipulated by the pleasantness-similarity between the two photos). In general, hard choices take longer to resolve than easy choices, a finding we report in the companion paper and replicate here in our analysis of Response Times (see Figure 3.1). However, response time is a coarse measure that cannot reveal any of the underlying decision processes. Similarly, though it is able to fill in the gap about the effects of decision-difficulty on decision processes *after* a movement is initiated, mouse-tracking (e.g., Ouellette Zuk et al., 2023) is blind to decision processes arising *prior* to movement onset. Thus, the

Measure	M Decision Difficulty Hard	Easy	Hard-Easy *M*	*SE*	df	*t*	*p*	Cohen's *d*
Sentence Verification								
1 Dwell on Chosen								
First Dwell Onset (ms)	1857	1149	708	41.5	88	17.0	***	1.81
First Dwell Duration (ms)	227	214	13	11.8	88	1.1	n.s.	0.12
End Dwell to Response (ms)	89	42	46	14.5	88	3.2	*	0.34
Numeric-Size Congruity								
1 Dwell on Chosen								
First Dwell Onset (ms)	448	429	19	9.0	89	2.1	n.s.	0.22
First Dwell Duration (ms)	302	267	35	8.2	89	4.3	**	0.45
End Dwell to Response (ms)	74	83	-9	6.5	89	-1.4	n.s.	-0.15
2 Dwells with the Last Dwell on Chosen								
First Dwell Onset (ms)	263	263	-0.5	8.3	87	-0.1	n.s.	-0.007
First Dwell Duration (ms)	207	217	-10	4.9	87	-2.0	n.s.	-0.21
Transition Duration (ms)	150	140	10	6.4	87	1.6	n.s.	0.17
Last Dwell Duration (ms)	248	211	37	8.9	87	4.2	**	0.45
End Dwell to Response (ms)	36	29	7	3.3	87	2.0	n.s.	0.21
Photo Preference								
2 Dwells with the First Dwell on Left, and the Last Dwell on the Right Chosen								
First Dwell Onset (ms)	282	272	10	17.7	79	0.6	n.s.	0.065
First Dwell Duration (ms)	524	494	30	48.6	79	0.6	n.s.	0.069
Transition Duration (ms)	156	131	25	10.7	79	2.4	n.s.	0.27
Last Dwell Duration (ms)	719	514	205	35.6	79	5.7	***	0.64
End Dwell to Response (ms)	51	39	12	8.9	79	1.3	n.s.	0.15
2 or More Dwells with the Last Dwell on Chosen								
First Dwell Onset (ms)	295	282	13	9.5	94	1.4	n.s.	0.14
First Dwell Duration (ms)	454	477	-23	13.2	94	-1.8	n.s.	-0.18
Last Dwell Onset (ms)	1482	1261	221	31.2	94	7.1	***	0.73
Last Dwell Duration (ms)	521	407	114	19.1	94	6.0	***	0.61
End Dwell to Response (ms)	117	59	57	14.1	94	4.1	**	0.42

Table 3.2: Pairwise comparisons between hard and easy trials for all gaze metrics. Note. Gaze dynamics were tested at an alpha level of .05/21 = .0023809. *p < .0023809; **p < .0001; ***p < .00001

primary objective of this study was to examine how decision difficulty manifests in gaze dynamics - a measure capable of indexing decision processes from the moment of stimulus onset.

In order to investigate gaze dynamics it was first necessary to more broadly categorize gaze patterns - the series of looks (dwells) the eyes made on certain targets relevant to a decision. For example, it would be meaningless to examine the dynamics of a second-dwell for a task where this rarely occurred. We first examined trial characteristics including number of dwells, the side of space of the first dwell and whether the last dwell was toward the chosen target, calculating the proportion of trials observed for each gaze pattern and whether they occurred in hard or easy trials. This analysis of common patterns first revealed a task-general difficulty effect whereby more dwells were observed on harder trials. Second, and more importantly, it also highlighted distinct and unique gaze patterns observed *between* tasks. While the first paper in this series (Ouellette Zuk et al., 2023) demonstrated the consistency of decision difficulty effects between tasks, here the proportion of trials analysis exposed a relationship between gaze behaviour and the spatial distribution of decision information within a task.

In the Sentence Verification task all of the decision information is contained in a statement at the top-middle of the screen with no unique information contained at the left or right choice options (the "True" and "False" labels at these locations remained constant). As such, the dominant gaze behaviour in this task contained either no looks toward the choice options or a single look toward the option that was selected. On the trials with one dwell on the chosen option, we found that the duration of that dwell remained the same for both hard and easy trials. Instead, the effects of decision difficulty emerged in the time it took for the dwell to start, and the time from the end of the dwell to the response. In this task we can infer that the gaze was focused on the sentence, and as the sentence contained all of the decision information, the differentiation between hard and easy trials emerged prior to any look towards the response options. In the case where a subsequent look to the chosen target occurs, the constant dwell duration seen regardless of decision difficulty suggests that this gaze might only serve the difficulty-independent process of spatially guiding the mouse response. However, after the gaze leaves the chosen option, difficulty effects re-emerge, suggesting that additional cognitive processing may be required beyond the choice-option dwell to finalize the decision.

In the Numeric-Size Congruity task, unlike Sentence Verification, the information necessary to make a decision is contained at the choice options. However, this task has the interesting property that sometimes a single fixation toward a choice option is sufficient to make a decision (e.g., if your eyes dwell on the digit “1” or “9” you can definitively know it is the lower or higher numeric value respectively) while other times a single dwell is insufficient (e.g., if you eyes dwell on the digit “2” or “8” fixating the other target is necessary to make a definitive decision). Given this, it is logical that the dominant gaze patterns observed in this task are divided into trials with only a single dwell and trials with two dwells. When examining the single dwell on chosen trials of the Numeric-Size Congruity task, difficulty effects manifest in a manner wholly opposite to that of the Sentence Verification task. Specifically, when participants focus only on the chosen option, they spend significantly more time dwelling on it during hard trials compared to easy trials. This prolonged dwell on the incongruent yet correct choice option is the only metric in the Numeric-Size Congruity task where difficulty effects become evident. Neither the onset of the first dwell nor the offset to response show any changes with difficulty. This begins to fill in the picture of how eye gaze functions with respect to decision difficulty - at the moment when sufficient information about the decision has been acquired via eye gaze then, and only then, does difficulty begin to differentially affect the decision timeline.

This hypothesis of gaze distribution being tied to information acquisition receives additional support from the two-dwell trials in the Numeric-Size Congruity task. Here we observe an intriguing pattern whereby the gaze dynamics of hard and easy trials appear identical all the way until the second dwell, where then, and only then, do hard trials exhibit a significantly longer duration for the last dwell compared to easy trials. Following this extended second dwell, the time to the response is comparable for both difficulty conditions. Thus, all the effects of decision difficulty for this set of trials are reflected exclusively in the last dwell, where determining the correctness of a numerically large but physically small (i.e., incongruent) choice option requires more dwell time compared to a numerically large and physically large (i.e., congruent) choice option. Perhaps most importantly, the first dwell, which is predominantly directed towards the unchosen option, does not take longer in the hard condition. It is only when participants view both choice options, and therefore have acquired all the necessary decision information, that difficulty effects emerge.

Finally in the Photo Preference task, decision information is necessarily and evenly distributed between both choice targets. That is, you can't make a determination of comparative preference between two photos without having your eyes dwell on both of them. Accordingly, when examining the most common gaze patterns in the Photo Preference task, they all involve two or more dwells with the single most common pattern being a look to the left, unchosen target followed by a look toward the right, chosen target. Focusing on this specific two-dwell pattern, the results follow the structure outlined in the Numeric-Size Congruity task. That is, the difficulty effect emerges exclusively in the last dwell duration of these Photo Preference trials. Like before, a hard trial does not exhibit signs of being difficult until the second, chosen option is viewed.

These results are confirmed in our broader investigation of Photo Preference trials with two or more dwells (e.g., trials with two, three, four or more dwells). Again, we consistently observed difficulty effects in the latter half of the trial, with hard trials showing a longer duration for the last dwell, as well as later onsets of the last dwell and longer durations between the last dwell and the response. This broader result shows the consistency of the difficulty effect on the duration of the last dwell, while the prolonged onset of the last dwell on hard trials likely results from averaging trials with a higher dwell count. Of note, the time from the offset of the last dwell to the response should not be conflated with the number of dwells in the same way. Instead, it suggests that this broader set of trials, with more dwells in harder trials, may more clearly capture decisions with lingering uncertainty. In other words, if harder trials take longer and more dwells cost more time, this set of trials may reflect the more challenging choices where difficult decisions are still being resolved all the way until the final choice occurs.

Taken together, the analysis of gaze behaviour across these three binary choice tasks sheds light on the dynamics of decision making. It appears there are at least two processes at play - the gathering of decision information and the resolution of the decision. While these two aspects of the decision can likely proceed in parallel, our data suggest that information gathering is the predominant driver of gaze early in trials and proceeds largely without impact from the specific demands of decision difficulty until all the relevant information has been at least partially sampled. Moreover, information gathering appears to be highly stereotyped for a given task. This is most evident in the Photo Preference trials - when information was evenly distributed across two locations, the vast majority of dwells were directed first to the left, and then to the right. This aligns with gaze-focused decision models trained on behavioral data from similar binary choice preference tasks (Busemeyer & Townsend, 1993; Krajbich et al., 2010; Shimojo et al., 2003). These models assume a left-first gaze, highlighting the persistent influence of ingrained eye movements used in reading left to right (at least in the English-speaking population tested). Additional evidence for the stereotyped nature of information gathering is seen in the first dwell dynamics on trials with more than one dwell (Numeric-Size Congruity and Photo Preference). Here, the time to first dwell and first dwell durations are not impacted by decision difficulty.

This suggests that there is a level of dynamics involved in decision making that most models don't capture. That is, most decision-models simulate a dynamic decision using a series of static parameters - for example the rate at which evidence for an option accumulates and the bound to which it must accumulate to in order for it to be chosen. However, here we show that the parameters themselves are likely changing throughout a decision. Specifically, we argue that the value of a given choice option necessarily fluctuates as the decision shifts from information gathering to decision resolution. Consider the Photo Preference task. Initially a target's value is dictated by its ability to deliver new information (for a review of this idea, see Gottlieb et al., 2013), separate from its content (e.g., pleasantness). Thus, both targets are equally valuable and the eyes adopt a stereotyped left-to-right pattern of information sampling. But then, target-value shifts to being defined by the details of the image. Now, the pleasantness of each option, and critically the relative pleasantness between options, dictates the resolution of the decision. This ability to shift decision parameters based on the current task context can explain both of our major gaze dynamic findings: initial dwells are stereotypical, driven by information gathering and not affected by decision difficulty, while the last dwells are affected by decision difficulty since value during decision resolution is determined by the relative difference in task-relevant content between targets.

Understood this way, it is clear that multiple facets of a target determine its value: if it contains task-relevant information, if it has been looked at, if it is an image or text, if it is small or large, if it is easily identified or not. Equally clear is that which facets are of value changes over time. Across our two studies we can measure the value transition from information gathering to decision resolution - a specific example of the more general pattern of explore versus exploit behaviors (J. D. Cohen et al., 2007). Since a key aspect of exploration is the physical location of targets, eye-tracking is particularly well suited to measure these kinds of information gathering behaviours. Then, as a decision shifts to resolution, which typically demands a motor response, mouse-tracking becomes a sensitive tool for watching the later stages of competition play out across time and space. Together, this highlights the complementary approaches taken across our two companion papers: gaze behaviour is acutely sensitive to task differences, especially early in trials and with respect to the spatial distribution of decision-relevant information, while mouse tracking is more sensitive to the decision difficulty effects that appear once all relevant information has been sampled.

We believe a cross-paper comparison of difficulty effects between tasks offers some initial support for this idea. Arguably, the gaze dynamics during Sentence Verification are the least informative - many trials have no dwells at the site of a choice and those that do, don't have dwell durations that differentiate between hard and easy trials. In complete contrast, as we report in our companion paper (Ouellette Zuk et al., 2023), during-movement measures of movement time and trajectory strongly differentiate between hard and easy trials for the same task. On the other end of the spectrum, the current analysis of gaze patterns in the Photo Preference task offers rich information about the contemplation of decision options, including many trials with multiple dwells. But, this same task in our companion paper (Ouellette Zuk et al., 2023) shows that during-movement measures had, relative to the other tasks, the least sensitivity to decision-difficulty. Not only does this show why collecting gaze and movement data is important, it also has theoretical implications. We previously made the distinction between gathering and resolving decision information - related processes that can sometimes proceed in parallel. We speculate that tasks like Photo Preference which require longer times spent gathering information are thereby also granted extra time to start resolving a decision prior to movement onset. As a result, movement related measures in these kinds of tasks show less sensitivity as more of the decision has been completed prior to the initiation of a response.

Aside from their impressive combined ability to cover the full range of a decision - from stimulus onset to response completion - there is another, methodological link between our two companion papers: their use of remote data collection. Not only did it vastly increase the sample size of our studies (more than 400 data sets initially collected across the two papers) it also made the study accessible to participants who may not typically participate in academic research (see supplementary material in Ouellette Zuk et al., 2023). Maybe most importantly, it also allowed us to test participants in more ecologically-valid environments. That is, we can collect data from people using their own devices from the comfort of their own homes without introducing an artificial, isolated, and highly-controlled laboratory setting that likely limits our ability to capture realistic and natural human behaviour (Kingstone et al., 2008; Shamay-Tsoory & Mendelsohn, 2019).

3.4 Limitations and Conclusion

Together, this paper series demonstrates the power of remote, online methods as tools for deeply understanding the complete, dynamic and continuous decision process, rewinding the decision from the typically-collected final choice response, all the way back to the first glance. Combined, our studies reveal an incredibly robust set of findings, on smartphones, tablets and computers with mouse-tracking or eye-tracking across three unique decision tasks we can measure and decompose the effects of decision difficulty with precision.

Of course, full remote data collection is not without its limitations, some which are particularly evident in the current study. Most notably, there are known spatial and temporal inaccuracies when using webcam eye-tracking. As an example, we did not feel we had sufficient spatial accuracy to properly analyze nuanced reading behaviour during the initial dwells to the statement in the Sentence Verification task. This limitation also led to an unforeseen and unfortunate outcome - due to the sampling rate slowdown caused by prioritizing the collection of webcam eye-tracking, our mouse-tracking data in the current study was not sufficiently sampled to perform a confirmatory analysis of the effects reported in our companion paper (Ouellette Zuk et al., 2023). Finally, at present, webcam eye-tracking is generally restricted to participants using a computer with a webcam - it hasn't yet been widely used with sufficient accuracy on mobile devices in decision science research.

These limitations, however, are the type which seem likely to be overcome soon. Improved and more efficient gaze detection driven by ever-improving machine learning models, continued advancements in consumer hardware, and the use of mobile device cameras for eye-tracking research (e.g., Namnakani et al., 2023) are all on the imminent horizon (Khamis et al., 2018). And in that future, we envision using gaze and movement tracking will be paramount to understanding participant behaviour in and out of the lab. Importantly, as we advocate for moving into unrestricted domains - where the nature of the task and the decisions that people make are not controlled by an experimenter - we recognize the need for and strength of this combined metrics approach. Including both gaze and movement analysis allows you to understand both where the most relevant decision information is located via eye-tracking, and via motion-tracking, how difficult it is to adjudicate between that information to arrive at and perform the final movement required to enact a choice.

3.5 Methods

3.5.1 Participants

100 adults provided their informed consent to participate in the experiment, and completed the study in full. Of the 100 participants, 36 self-identified as female, 66 as male, and one participant preferred to not disclose their gender. The average participant age was 25.47 years old (+/- 4.28). Participants were recruited using Prolific (www.prolific.co), an online crowdsourcing platform, where we followed Ouellette Zuk et al.'s (2023) participant restrictions on age (18 to 35 years old), and prior approval rating on the platform (95-100%). We paid participants for their time (6 GBP per hour, ~$10 CAD per hour). All experimental proceedings were approved by the University of Alberta's Research Ethics Board (Pro00087329) and were performed in accordance with relevant guidelines and regulations.

3.5.2 Materials

All participant data for the study was collected through the use of Labvanced (Finger et al., 2017), an online, browser-based Javascript experimentation platform. We designed our study in a 800 x 450 pixel coordinate frame in Labvanced (see Figure 3.3B), where it would automatically scale to the size of the participant's screen. We used Labvanced's built-in webcam eye-tracking (Labvanced v2 High Sampling Mode eye-tracking Finger et al., 2017). To ensure high-quality data collection, certain minimum requirements were imposed. Participants were required to use either a laptop (n = 75) or desktop (n = 25) computer with a mouse. The operating systems supported were Mac (n = 15), Windows (n = 84) or Linux (n = 1), along with the Chrome browser. Furthermore, participants were required to have a webcam with a minimum resolution of 1280 x 720 pixels, a landscape-oriented screen with a minimum resolution of 600 x 600 pixels (*Mode* = 864 x 1536 px), and a computer system capable of collecting at least 10 samples per second of the head's position for optimal eye-tracking precision (M = 15.65 Hz, SD = 5.69 Hz). This system threshold was often met if participants had a graphics card and had freed up system resources prior to starting the study (i.e. closing any other programs running on their computer).

3.5.3 Task & Procedure

Broadly, the tasks and procedures followed by Ouellette Zuk et al. (2023) were repeated here but with the addition of webcam eye-tracking, and with only computers being included (Ouellette Zuk et al also tested tablets and smartphones). Like Ouellette Zuk et al. (2023), we asked participants to complete three distinct tasks that all required decision-making in a binary choice paradigm where choices were made with mouse movements. These tasks were Numeric-Size Congruity, Sentence Verification, and Photo Preference (with examples shown in Figure 3.1).

Participants recruited through Prolific (www.prolific.co) were given access to a detailed description of the study. This description included an approximate duration of the study (1 hour), information regarding the necessary hardware, and instructions on how to avoid any potential technical difficulties (complete Prolific description available in the Supplementary Materials). Upon clicking the study link that accompanied the study's description, participants were directed to a full-screen Labvanced browser window and prompted to grant permission for their webcam device. In the event that participants did not meet the minimum requirements, they would immediately receive an error or warning message. Assuming no issues arose, participants would begin by providing their informed consent to participate, after which they would proceed to answer a brief survey pertaining to their demographic information and the hardware they were using.

Before the primary experimental tasks, participants were given information to encourage successful webcam eye-tracking data collection. Instructions about optimal lighting conditions, the re-calibration process, and the virtual chinrest feature were provided to participants. Then, participants underwent Labvanced's 5-minute eye-tracking calibration procedure. It was required that participants redo the calibration if the predicted gaze error surpassed 7% of the screen's dimensions. After completing calibration, participants began the main experiment.

Figure 3.3 illustrates the experimental procedure and trial progression. Three tasks were performed, with simple task instructions preceding each task. The presentation order of the tasks was counterbalanced such that participants were automatically assigned to one of the six possible order variations by the online experimentation platform in a manner that tried to balance the number of completed datasets across all six order variations. 84 trials were completed per task, with task stimuli presented in a randomized order within each task. Every 5 trials, a brief seven-point eye-tracking recalibration procedure was performed to adaptively correct any drift errors in the gaze prediction algorithm over the course of the experiment. At any point during an experimental trial, the virtual chinrest feature would pause the trial if a participant's excessive head movement affected the quality of the gaze prediction ($M = 6.59$ trials,

SD = 7.22 trials).

An experimental trial began with a green circular start button labeled "Touch here" at the bottom center of the screen. Participants had to move their mouse cursor to the button to initiate the trial. This prompted the appearance of a three-second countdown at the center of the screen. If the mouse cursor was removed from the circular button, the countdown paused until the cursor's return. In the Numeric Size-Congruity and Photo Preference tasks, a task-specific question appeared at the top center of the screen during the countdown (see Figure 3.1). Once the countdown ended, two choice boxes appeared at the upper-left and upper-right corners of the screen, presenting trial-specific options. In the Sentence Verification task, two choice options appeared alongside the countdown and displayed a statement at the top center of the screen after the countdown completed. Participants could immediately select their choice option by moving their mouse cursor inside the respective choice area. Once a choice was made, the selected box was highlighted while the other choice option and start button disappeared. A "Next" button then appeared at the center of the screen for participants to click to proceed to the next trial at their own pace.

All three tasks required binary choice decisions in Hard and Easy conditions. In the Numeric-Size Congruity task, participants were presented with pairs of digits and asked to determine which digit had a higher numeric value. The pairs of digits varied in congruence, where some pairs were congruent in both numeric and physical size (representing Easy trials with low decision difficulty, e.g., $\scriptstyle 2$ vs. $\large 8$), while others were incongruent in numeric and physical size (representing Hard trials with high decision difficulty, e.g., $\large 2$ vs. $\scriptstyle 8$). In the Sentence Verification task, participants were tasked with verifying the truthfulness of statements. From previous work (Dale & Duran, 2011) it has been shown that statements that are true show large decision difficulty effects based on whether they are non-negated (representing Easy trials with low decision difficulty, e.g., 'Cars have tires') or negated (representing Hard trials with high decision difficulty, e.g., 'Cars do not have wings'). In the Photo

Preference task, participants were presented with pairs of photos that differed in valence (from the International Affective Picture System stimulus set, Lang et al., 2008, as in Koop and Johnson, 2013). They were asked to then choose which photo they preferred. The pairs of photos varied in their dissimilarity of valence, with some pairs being dissimilar (representing Easy trials with low decision difficulty, e.g. High vs. Low pleasantness) and others being similar (representing Hard trials with high decision difficulty, e.g. High vs. High pleasantness). These tasks were designed to cover a wide range of decision domains, including objective perceptual judgments (such as discriminating between digits), semi-subjective conceptual judgments (such as evaluating the truth value of statements), and subjective preference judgments (such as expressing a preference for specific photographs). Additionally, the tasks intentionally differed in terms of stimulus characteristics and the cognitive processing requirements involved.

The entire experimental procedure, as a Labvanced study, can be accessed via the link in Supplementary Materials.

3.5.4 Data Processing

The uncontrolled nature of online, remote data collection, including the use of webcam eye-tracking, presented some data quality challenges that required thoughtful treatments. Gaze and cursor timeseries data were collected in a way that maximized the number of data samples processed for each participant, with priority given to the collection of gaze samples. All gaze and cursor data were then upsampled (linearly interpolated) to a common sampling rate of 60 Hz. The gaze prediction algorithm was refined (i.e., recalibrated) every 5 trials using Labvanced's adaptive drift correction method (Finger et al., 2017), therefore gaze data quality varied over time. To minimize data rejection, we assessed the quality of each participant's gaze data within each task independently (as opposed to rejecting an entire dataset for poor gaze data for a subset of trials).

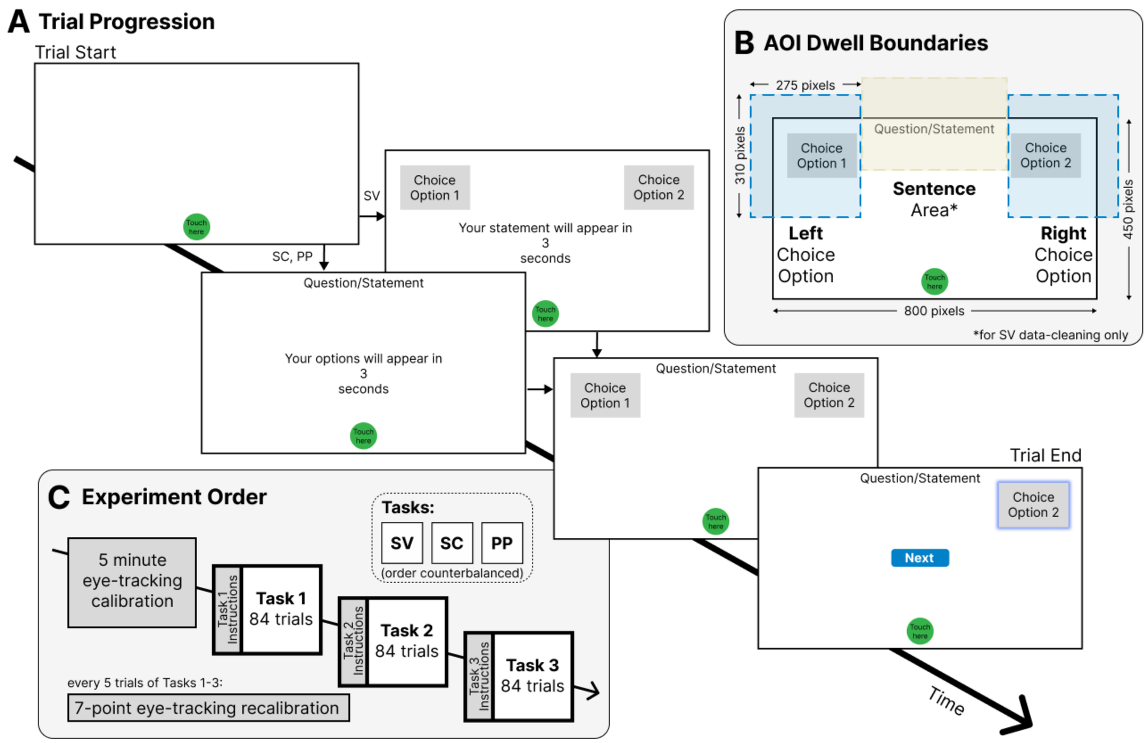


Figure 3.3: A) All three tasks presented a classic reach-decision paradigm requiring participants to choose one of two stimuli presented at the top left and right of their computer screen. For Numeric-Size Congruity (SC) and Photo Preference tasks (PP), countdown onset was accompanied by a question specific to the task, appearing at the top center of the display. The Sentence Verification (SV) task presented the two choice options coincident with countdown onset and presented a statement (rather than a question) upon countdown completion. Participants proceeded in a self-paced manner, pressing a button to begin the next trial. B) The enlarged areas of interest (AOIs) used to define the boundaries of the Left and Right choice options (blue transparent areas) when determining whether a dwell was made. The dimensions of the choice AOIs (275 x 310 pixels) are presented relative to the dimensions of the frame size (800 x 450 pixels). The Question/Statement AOI was used only for data-cleaning in the SV task. This frame was scaled to the size of each participant's screen. C) Overview of the experiment's design. Each participant completed an SV task, an SC task, and a PP task, with task order counterbalanced between participants. Eye-tracking calibration occurred at the beginning of the session, and a re-calibration procedure was performed every 5 trials. Task-specific instructions were presented prior to each task.

Using custom MATLAB scripts and our Gaze and Movement Analysis software, our initial data cleaning approach aimed to assess whether the gaze data showed reasonable patterns or whether it contained noisy, spurious gaze prediction errors (see Bertrand and Chapman, 2023 for a similar approach). Using more exaggerated but still mutually-exclusive boundaries (see Figure 3.3B - AOI Dwell Boundaries), we determined whether the gaze fell inside at least one of the task-critical areas during the decision period. For the Numeric-Size Congruity and Photo Preference tasks, we considered the gaze data of reasonable quality if the gaze fell within the left or right choice options for at least 100 milliseconds continuously. For the Sentence Verification task, we instead used at least one look (minimum 100 ms) to the sentence area as a proxy of reasonable gaze data as it was common for participants to not look at the choice options. In each task, if less than half of the trials showed reasonable gaze data (i.e., one dwell on a task-relevant AOI), we removed the full task from the participant's dataset for analysis. Based on this criteria, from 100 participants, 3 subjects' Photo Preference task data were removed, 9 subjects' Numeric-Size Congruity task data were removed, and 3 subjects' Sentence Verification task data were removed. Of the remaining datasets, any individual trials that failed to show reasonable gaze data (per the same criteria) were removed leaving the Photo Preference, Numeric-Size Congruity, and Sentence Verification tasks with 82.79 (+/-2.96), 73.41 (+/-9.58) and 81.70 (+/-5.79) of 84 trials respectively.

Then, to assess non-gaze-related quality of data, we employed similar rejection criteria to Ouellette Zuk et al. (2023). Within each task, we removed any trials where the response time was less than 100 milliseconds or greater than 3 standard deviations above the subject's mean response time in that task. We also removed any trials where a pause occurred from the virtual chinrest feature, or any trials where the response time was not computable (from participant error or occasional data recording issues). Further, any incorrect trials were removed from the Sentence Verification and Numeric-Size Congruity tasks (where accuracy could be assessed objectively).

Following these additional trial rejections, we again removed entire task data from any participant with less than 42 of their original 84 task trials, which resulted in one additional Numeric-Size Congruity task removal. From 100 participants, the final task datasets included in analysis were: Photo Preference: n = 97, 78.30 (+/- 5.18) trials, Numeric-Size Congruity: n=90, 70.87 (+/-9.16) trials, Sentence Verification: n=97, 72.09 (+/-8.23) trials.

Unlike our companion paper (Ouellette Zuk et al. 2023) where cursor-tracking provided the key dependent measures, high-quality gaze data from webcam eye-tracking was the primary goal of the current study. While we did record mouse trajectory data alongside our gaze data, we did not have sufficient data quality to reliably metricize the mouse trajectory data in the same way as Ouellette Zuk et al. (2023). With recording priority given to the gaze data, and cursor movements generally being very quick, we found that there were many instances that the number of data samples from the cursor were insufficient. We discuss this unintended consequence in the Discussion, and encourage readers to engage with our companion paper (Ouellette Zuk et al., 2023), which provides a thorough analysis of high-quality cursor and touchscreen trajectories from an original sample of more than 300 other participants.

3.5.5 Dependent Measures

Our analysis strategy revolved around three sequential steps. First, to confirm and replicate that the main decision-difficulty effects elicited by these tasks were present in the current study, for every trial we recorded *Response Time (ms)*: the time from the choice options being presented to the moment the cursor was detected as entering within the bounds of a choice option.

Second, to broadly characterize and analyze the dominant gaze patterns within each task, we calculated *Proportion of Trials (%)*: a value of 0 or 1 for each trial that represented if that trial shared certain characteristics (e.g. was it Hard or Easy). These counts were then aggregated such that proportion of trials for a given characteristic was always calculated within each task per participant, where the trials fitting the characteristic were counted and then divided by the total number of trials included in the analysis (specifically, the number of Hard plus Easy trials).

Third, we used the results from the proportion analysis to guide our examination of the gaze dynamics of the most frequent gaze patterns within each task. These gaze dynamics were centered on describing the timing of specific dwell patterns - when a dwell started, how long it lasted, and when it ended relative to the response. For patterns with more than one dwell, we wanted to capture this information about both the first and last dwell. Using the first and last (as opposed to strictly second) dwell afforded us flexibility in describing two-dwell patterns, but also three, four or more dwell patterns. In all three tasks, for every gaze pattern observed (where every gaze pattern necessitated there being at least one dwell), the following measures were collected:

First Dwell Onset (ms): the time from the choice options being revealed to the start of the first dwell within a choice option.

First Dwell Duration (ms): the length of time the dwell stayed at the first choice option viewed.

End Dwell to Response (ms): the time from the end of the final dwell to the response, as determined by the mouse cursor's entry into the choice option.

When a frequent gaze pattern included more than one fixation (only the Numeric-Size Congruity and Photo Preference tasks), the last dwell's duration was also collected:

Last Dwell Duration (ms): the length of time the dwell stayed at the last choice option viewed preceding a response.

To capture the time of the last dwell's onset, we used two measures, dependent on the number of dwells in the gaze pattern being explored. When looking at trials with exactly two dwells, we use Transition Duration as a measure, but when looking at a collection of trials with two or more dwells (Photo Preference only), we use Last

Dwell Onset:

Transition Duration (ms): the time between the first and second (i.e. last, in these cases) dwells, measured from the offset of the first dwell to the onset of the second dwell.

Last Dwell Onset (ms): the time from the choice options being revealed to the start of the last (i.e. second, third, fourth or more) dwell within a choice option.

3.5.6 Statistical Procedure

Each dependent measure was analyzed within each task using Jamovi (Version 2.2.5; an open-source statistical software). Response times and gaze dynamic metrics for each task were analyzed using paired t-tests of each participant's Hard and Easy trial means on that given task. To correct for multiple comparisons, the three response time tests were performed with an alpha of $.05/3 = .01667$, and the twenty-one gaze dynamic metrics were tested with an alpha of $.05/21 = .00238$. The proportion of trials measures were tested using Repeated Measures ANOVAs, where p-values were Greenhouse-Geisser-corrected for sphericity violations. The Sentence Verification task was tested with a 3 factor RMANOVA, and Numeric-Size Congruity and Photo Preference tasks were tested with 4 factor RMANOVAs. We followed the family-wise error correction procedure from Cramer et al. (2016), where the threshold for significance becomes increasingly more conservative with every significant test result within a family of results. We treated all 3 omnibus RMANOVAs as a single family to determine the significance of the omnibus results for the proportion of trials measures. Follow-up RMANOVAs were then performed on the highest order interaction(s), testing each level of one factor against the other factors (see section 3.2 - Results). This interaction procedure was performed as necessary until a single-factor RMANOVA was reached, where the simple main effects of one factor could be tested at all levels of the other factor. Significant main effects were explored with all pairwise comparisons. The Cramer et al. (2016) procedure was again employed for these follow-up

RMANOVAs, where the family-wise error correction was performed within-task (i.e. each task's follow-up tests became a family). We report our results in Tables 3.1 and 3.2. Table 3.1 presents the per task trial proportion results from the breakdown of each three-way interaction involving the factor of difficulty, and Table 3.2 fully reports the gaze metrics differences between the Hard and Easy trials for each of our 21 gaze-dynamic measures.

Acknowledgments

This study was funded by an NSERC Discovery Grant to CSC, and a MITACS Accelerate International award to JKB.

3.6 References

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations [Publisher: SAGE Publications Inc]. *Journal of Management*, *47*(4), 823–837.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407.

Bánki, A., de Eccher, M., Falschlehner, L., Hoehl, S., & Markova, G. (2022). Comparing online webcam- and laboratory-based eye-tracking for the assessment of infants' audio-visual synchrony perception. *Frontiers in Psychology*, *12*, 733933.

Bertrand, J. K., & Chapman, C. S. (2023). Dynamics of eye-hand coordination are flexibly preserved in eye-cursor coordination during an online, digital, object interaction task. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–13.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment [Place: US Publisher: American Psychological Association]. *Psychological Review*, *100*, 432–459.

Cassey, T. C., Evens, D. R., Bogacz, R., Marshall, J. A. R., & Ludwig, C. J. H. (2013). Adaptive sampling of information in perceptual decision-making [Publisher: Public Library of Science]. *PLOS ONE*, *8*(11), e78993.

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, *33*(1), 269–298.

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should i stay or should i go? how the human brain manages the trade-off between exploitation and exploration [Publisher: Royal Society]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933–942.

Cramer, A. O. J., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R. P. P. P., Waldorp, L. J., & Wagenmakers, E.-J. (2016). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic Bulletin & Review*, *23*(2), 640–647.

Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science*, *35*(5), 983–996.

Dotan, D., Pinheiro-Chagas, P., Al Roumi, F., & Dehaene, S. (2019). Track it to crack it: Dissecting processing stages with finger tracking. *Trends in Cognitive Sciences*, *23*(12), 1058–1070.

Faulkenberry, T. J., Cruise, A., Lavro, D., & Shaki, S. (2016). Response trajectories capture the continuous dynamics of the size congruity effect. *Acta Psychologica*, *163*, 114–123.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: A unified JavaScript framework for online studies. *International Conference on Computational Social Science*, *1*(1), 1–3.

Freeman, J. B. (2018). Doing psychological science by hand [Publisher: SAGE Publications Inc]. *Current Directions in Psychological Science*, *27*(5), 315–323.

Gallivan, J. P., Chapman, C. S., Wolpert, D. M., & Flanagan, J. R. (2018). Decision-making in sensorimotor control [Number: 9 Publisher: Nature Publishing Group]. *Nature Reviews Neuroscience*, *19*(9), 519–534.

Glaholt, M. G., & Reingold, E. M. (2009). The time course of gaze bias in visual decision tasks. *Visual Cognition*, *17*(8), 1228–1243.

Glaholt, M. G., Wu, M.-C., & Reingold, E. M. (2009). Predicting preference from fixations [Place: Italy Publisher: PsychNology Journal]. *PsychNology Journal*, *7*, 141–158.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making [_eprint: https://doi.org/10.1146/annurev.neuro.29.051605.113038]. *Annual Review of Neuroscience*, *30*(1), 535–574.

Gottlieb, J. (2018). Understanding active sampling strategies: Empirical approaches and implications for attention and decision research. *Cortex*, *102*, 150–160.

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information seeking, curiosity and attention: Computational and neural mechanisms. *Trends in cognitive sciences*, *17*(11), 585–593.

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, *18*(3), 384–401.

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). Lab.js: A free, open, online study builder. *Behavior Research Methods*, *54*(2), 556–573.

Johnson, B. P., Dayan, E., Censor, N., & Cohen, L. G. (2021). Crowdsourcing in cognitive and systems neuroscience [Publisher: SAGE Publications Inc STM]. *The Neuroscientist*, 10738584211017018.

Khamis, M., Alt, F., & Bulling, A. (2018). The past, present, and future of gaze-enabled handheld mobile devices: Survey and lessons learned. *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–17.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317–340.

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, *67*(4), 151–185.

Krajbich, I. (2019). Accounting for attention in sequential sampling models of decision making. *Current Opinion in Psychology*, *29*, 6–11.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology*, *3*. Retrieved May 25, 2023, from https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00193

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Pro-*

ceedings of the National Academy of Sciences of the United States of America, *108*(33), 13852–13857.

Lang, P., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS) : Affective ratings of pictures and instruction manual [Publisher: University of Florida]. *Technical Report*. Retrieved May 31, 2023, from https://cir.nii.ac.jp/crid/1573950399053852928

Leeuw, J. R. d., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351.

Moher, J., & Song, J.-H. (2014). Perceptual decision processes flexibly adapt to avoid change-of-mind motor costs. *Journal of Vision*, *14*(8), 1.

Namnakani, O., Abdrabou, Y., Grizou, J., Esteves, A., & Khamis, M. (2023). Comparing dwell time, pursuits and gaze gestures for gaze interaction on handheld mobile devices. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–17.

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*(5), 1.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions [Place: United Kingdom Publisher: Blackwell Publishing]. *Psychological Science*, *9*, 347–356.

Schouten, J. F., & Bekker, J. A. M. (1967). Reaction time and accuracy. *Acta Psychologica*, *27*, 143–153.

Schuetz, I., Murdison, T. S., MacKenzie, K. J., & Zannoli, M. (2019). An explanation of fitts' law-like performance in gaze-based selection tasks using a psychophysics approach. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13.

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451–465.

Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-life neuroscience: An ecological approach to brain and behavior research [Publisher: SAGE Publications Inc]. *Perspectives on Psychological Science*, *14*(5), 841–859.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317–1322.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*(2), 135–168.

Smith, S. M., & Krajbich, I. (2018). Attention and choice across domains. *Journal of Experimental Psychology: General*, *147*(12), 1810–1826.

Stewart, E., Valsecchi, M., & Schütz, A. C. (2020). A review of interactions between peripheral and foveal vision. *Journal of Vision*, *20*(12), 2.

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences*, *22*(6), 531–543.

Stone, S. A., & Chapman, C. S. (2023). Unconscious frustration: Dynamically assessing user experience using eye and mouse tracking. *Proceedings of the ACM on Human-Computer Interaction*, *7*, 168:1–168:17.

Wispinski, N. J., Gallivan, J. P., & Chapman, C. S. (2020). Models, movements, and minds: Bridging the gap between decision making and action. *Annals of the New York Academy of Sciences*, *1464*(1), 30–51.

Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, *16*(6), 1485–1505. Retrieved August 24, 2022, from https://journal.sjdm.org/21/210525/jdm210525.html

Chapter 4

A practical guide to webcam eye-tracking

4.1 Introduction

This document is born out of a pandemic pivot from the (inaccessible) lab to online research methods including webcam eye-tracking. We naively transitioned to the online eye-tracking space assuming this method would be relatively developed and rich with resources. However, unlike its lab counterpart, we quickly realized that online, webcam-based eye-tracking is still in its infancy - an experimental method at best. This became a major research focus for our lab. That is, even though our lab access has returned, we have spent considerable effort and exploration over the last number of years investigating the limits of webcam eye-tracking and what it can tell us about measuring human behaviour remotely.

This document reflects the countless methodological lessons learned as we traversed a large knowledge gap to generate high-quality webcam eye-tracking results (e.g. Bertrand and Chapman, 2023; Bertrand et al., 2023). Its central purpose is to share methodological knowledge as a means of demystifying and supporting the use of webcam eye-tracking in remote, online behavioural research. We hope this document can serve as a guide to the method of webcam eye-tracking, an offering of best practices, and a collection of how-to's, intended to give those with some lab-grade eye-tracking experience the information to confidently deploy their first remote webcam eye-tracking study. In particular, we have tried to share this information in a non-technical manner. As we've gone about developing our methods, we have opted for tools that defer the highly technical aspects of optimized gaze prediction algorithms and browser-based experiment deployment to experts. Thus, this document is meant for behavioural researchers like (but not limited to) linguists, economists, developmental psychologists, behavioural scientists, cognitive neuroscientists, and user experience and human computer interaction researchers rather than computer scientists or software engineers.

4.1.1 Background

Recording the movements of the eyes has come a long way since the creative yet cumbersome mechanical devices used over one hundred years ago (e.g. Delabarre, 1898). In this current century, specialized eye-tracking hardware (which the reader may be most familiar with) typically involves the use of infrared light cameras. These eye-trackers illuminate the eyes with invisible infrared light to capture reflections from the cornea to provide high-contrast video images. The changing appearance of the corneal reflection (or sometimes other features) in these images is then fed into the gaze prediction algorithm, which has often been trained with calibration data. These sophisticated infrared eye-tracking systems come in various forms, including desktop-mounted eye-trackers with cameras near a computer screen, and head-mounted eye-trackers with cameras directly on the participant's head. Generally speaking, these lab-based eye-tracking devices have proven highly effective in measuring eye movements and predicting the gaze's location, supporting an 'explosion' of eye-tracking experiments in the last 15 years (Carter & Luke, 2020).

However, certain limitations of lab-based infrared eye trackers have always been present, including their high cost, potential intrusiveness for users, and confinement to lab environments. In response, there has been a significant push towards developing more lightweight, portable, and cost-effective alternatives. Human-computer interaction researchers, in particular, have emphasized the need for such tools to use gaze as an input for computer use (Levine, 1981), which has the potential to be hugely transformative as an assistive technology. However, for some time, the main obstacle in creating lightweight and accessible eye-tracking solutions was the difficulty of obtaining high-quality eye images.

Fortunately, there have been significant technological advancements that have begun to tackle this challenge. Firstly, there has been a significant improvement in camera hardware, particularly with the miniaturization of high-quality web cameras that are now commonly integrated directly into computers. Alternative methods to gaze prediction algorithms have also been developed, specifically optimized for the lower contrast images captured under visible spectrum lighting conditions (e.g. Bäck, 2006). Sewell and Komogortsev (2010) were some of the first to show the promise of an unmodified, consumer-grade webcam as a tool for real-time eye gaze tracking, and further work on algorithms for real-time eye-tracking with laptop cameras continued (Meng & Zhao, 2017; Zheng & Usagawa, 2018). It was also around this time that toolboxes like Webgazer (Papoutsaki et al., 2016) and TurkerGaze (Xu et al., 2015) emerged as exciting, early webcam eye-tracking software programs, meant specifically to be deployed for remote, online research.

Since then, more sophisticated webcam eye-tracking options have emerged (e.g. RealEye, Wisiecka et al., 2022; Labvanced, Finger et al., 2017; OpenGaze, Zhang et al., 2019; see Shehu et al., 2021b for complete review of remote eye-tracking options), aided by more robust gaze predication algorithms that harness advancements in machine learning, like deep neural networks. For the most part, the novelty of these webcam eye-trackers has meant that most research involving their application has been focused on the more technical aspects of the method. Webcam eye-tracking's accuracy, validity, and performance has been assessed across a variety of domains including behavioral, psychological, and cognitive science (Bogdan et al., 2023; Schneegans et al., 2021; Semmelmann & Weigelt, 2018; Yang & Krajbich, 2021), online learning research (Hutt et al., 2023; Madsen et al., 2021; Robal et al., 2018; Zhao et al., 2017), linguistics (Slim & Hartsuiker, 2022; Vos et al., 2022), marketing (Schröter et al., 2021), and clinical optometry applications (Bruno et al., 2023). Additionally, feasibility studies have explored the potential of webcam eye-tracking for research with specific populations such as infants (Bánki et al., 2022), older adults with Alzheimer's Disease (Greenaway et al., 2021), and neurodiverse students in classrooms (Wong et al., 2023).

As evidenced by the small collection of recent, methodologically-focused works, webcam eye-tracking is still a relatively new method. These works have exposed several of the challenges that come with using this new method. The main limitations include issues with temporal and spatial accuracy, the complex implementation of algorithms with low-quality data, and the general challenges of online experimentation, such as limited participant oversight and technical difficulties. Unlike the well-established lab environment, webcam eye-tracking involves navigating unknown territory. However, our experience in overcoming these challenges has motivated us to create this guide. We recognize the tremendous potential of webcam eye-tracking as a method, and while it may require more effort from experimenters, we aim to demystify this exciting approach and enable you to conduct high-quality webcam eye-tracking experiments with ease.

4.1.2 Organization of this guide

With many things to consider across the entire experimental process, we have structured this guide in a fashion that will be familiar to experimentalists. We present information organized by when it is most relevant across the experimental process, with information best considered Before, During, or After Data Collection.

In the following Before Data Collection section, we provide a comprehensive overview of information and considerations that establish the groundwork for a well-executed webcam eye tracking experiment. We emphasize the active and hands-on role of the experimenter in understanding the iterative cycles of designing, building, and testing, despite the limitations and constraints posed by remote webcam eye tracking. We also discuss additional considerations prior to collecting data, like the timeseries nature of gaze data, and effective participant communication.

We then focus on the important considerations During Data Collection, taking evidence from our own experiments to frame the data collection process as a sort of filtering procedure. This data collection filter is described in both theory and practice, used to illustrate the ways that the collection process is affected by the experimenter's choices, and also a product of an uncontrollable remote data collection environment. We also discuss the cost implications of webcam eye-tracking as they relate to the data collection filter.

In the After Data Collection section, we assess the quality of the collected data in both the temporal and spatial domains. We explain the quality of our data within the context of our chosen system settings and the data filtering process. Additionally, we discuss data processing techniques, including those informed by features of the task, that can assist researchers in effectively addressing their research questions within the constraints of webcam eye-tracking.

4.2 Before Data Collection

We begin our guide with the Before Data Collection section, dedicated to the process of designing and constructing a high-quality webcam eye-tracking experiment. In this section, we outline the unique limitations and constraints of webcam eye-tracking and guide you through an iterative process that involves: A) designing an experiment that considers these constraints, B) exploring potential solutions for overcoming or working within these constraints, and C) rigorously testing your choices to ensure confidence in the experiment's quality. This section also covers essential but wide-ranging aspects like the collection of gaze data as timeseries data and participant communication, including recruitment and instructions. Although the process we will describe is not strictly linear (and rather is a recursive and iterative one), we present it in three distinct parts for the sake of clarity and document structure.

Before moving forward, we'd like to briefly acknowledge the importance of selecting a suitable webcam eye-tracking system to ensure a high-quality webcam eye-tracking experiment. Similar to laboratory eye-trackers, webcam eye-tracking software and web applications vary in their accessibility (e.g. open or closed sourced software), flexibility, user-friendliness, intended use, platform compatibility, data analysis tools, support, pricing, and, crucially, data quality. We have intentionally avoided providing an extensive review of webcam eye-tracking systems in this guide, but point the reader to the existing works that elaborate on features and functionalities of different webcam eye-tracking systems (e.g. Bánki et al., 2022; Shehu et al., 2021b; Vos et al., 2022). In our own search for webcam eye-tracking options for remote online testing during the global pandemic, we opted for Labvanced's webcam eye-tracking as it was offered within a complete online experimentation platform (something we also required as part of our pandemic transition). Moreover, following the iterative process we will next describe, our initial internal testing confirmed satisfactory eye-tracking data quality.

4.2.1 Experimental design within the constraints of webcam eye-tracking

The very first step to building a great webcam eye-tracking experiment is to gain a full appreciation for the challenges inherent to the method. The most notable difference for lab-based eye-tracking experimenters will be the significant limits to the spatial and temporal quality of webcam eye-tracking data. These limitations stem from differences in sampling rates between consumer-grade and lab-grade hardware (webcam vs. commercial eye-trackers) and the unpredictable and uncontrollable environment encountered during remote participant engagement. As a result, we must recognize that our representation of gaze data will exhibit more noise and deviation from the true gaze position when compared to data collected in controlled lab contexts.

Spatial noise arises due to the inaccuracy inherent to a gaze *prediction* algorithm, where it's not possible to achieve perfect accuracy in pinpointing the gaze's location. In and out of the lab, eye-tracking typically involves an initial calibration procedure and periodic re-calibrations to maintain a stable prediction of the gaze position. In the case of lab-grade systems like those from SR Research (Ottawa, Canada), Tobii (Stockholm, Sweden), or Pupil Labs (Berlin, Germany), researchers have grown accustomed to spatial errors on the order of less than one visual degree (Ehinger et al., 2019). Relying on predictive algorithms makes some degree of spatial error inevitable, but in webcam eye-tracking, the magnitude is greater. Reports on this are varied and limited, with estimates ranging from 0.88° (Skovsgaard et al., 2011) to 2.6° (Saxena et al., 2022) to ∼4° (Semmelmann & Weigelt, 2018) and even exceeding 7° in certain cases (Shehu et al., 2021a, in a comparison of various remote gaze estimation methods). Moreover, webcam eye-tracking's spatial accuracy can vary across different areas of the screen (Semmelmann & Weigelt, 2018; Vos et al., 2022).

Noise in the temporal domain arises due to limitations in the number of data samples collected over time. In laboratory settings, high-end hardware is designed to maximize the sampling rate, allowing specialized cameras to track the eye's position at rates from hundreds to thousands of times per second (e.g. the EyeLink 1000 can achieve up to 2000 Hz). This high sampling rate enables precise knowledge of when the gaze transitions from one location to another with millisecond precision. This allows researchers to ask questions not only about where they eye fixates, but also about how the eye moves (e.g. saccades and microsaccades). On the other hand, webcam eye-tracking systems rely on consumer-grade hardware. Currently, webcams typically offer a sampling rate of only 30 Hz, although 60 Hz webcams are gradually becoming more popular. Further, the samples collected by webcams (image data) need to be processed into timestamped gaze coordinates (positional data), a task usually performed using a participant's consumer-grade computing resources.

The temporal accuracy of webcam eye-tracking systems has been reported in limited contexts, with varying ranges: including 11.5 Hz (Bánki et al., 2022), 14 Hz (the remote cohort in Semmelmann and Weigelt, 2018), 15 Hz (Stone & Chapman, 2023), ∼20 Hz (Vos et al., 2022) and ∼40 Hz (Yang & Krajbich, 2021). Consequently, instead of recording gaze samples every half of a millisecond as can be achieved in the lab, webcam eye-tracking systems only produce an estimated gaze location at best every 16 or 32 milliseconds (depending on a respective 60 or 30 Hz hardware-limited rate), but in practice are often even slower (e.g. ∼47 ms on average based on our datasets presented in the During Data Collection section).

So, what implications do these less-precise spatial and temporal features of webcam eye-tracking data hold for aspiring researchers in the field? Primarily, the limitations of webcam eye-tracking restrict researchers' ability to address certain research questions or employ specific experimental designs that rely on spatial or temporal precision. This may seem obvious, but the limited spatial accuracy of webcam eye-tracking data makes it difficult to investigate research questions about the precise location of the gaze. This becomes particularly problematic in densely populated task spaces where multiple task-relevant objects are present, as it becomes difficult to discern precisely which object is being fixated. Consequently, answering questions about the gaze patterns during crowded visual search scenes or the reading of dense text passages may not be feasible with webcam eye-tracking. Recent studies have provided examples of the lower spatial accuracy limits of webcam eye-tracking systems, confirming the ability to, at a minimum, distinguish between 4 (Bertrand & Chapman, 2023; Slim & Hartsuiker, 2022) to 6 (Yang & Krajbich, 2021) areas of interest on a computer screen. In parallel, the temporal limitations of webcam eye-tracking impede experimenters from drawing conclusions about the precise timing and dynamics of gaze patterns, such as the dynamics of saccades. These limitations also impact the ability to investigate questions involving very small timing effects.

By now, it should be evident that webcam eye-tracking has inherent constraints on its utility, and researchers must consider these limitations when designing their experiments. The main questions that arise are: how can we practically account for these limitations and, how can we determine the threshold for crowding in a visual search display or define what qualifies as a 'too brief' looking behaviour?

In reality, there are no hard rules when it comes to addressing these types of questions regarding experimental design. Instead, experimenters must actively engage in a highly iterative process of understanding how the limitations of webcam eye-tracking can be considered and potentially controlled within their chosen webcam eye-tracking system (and are perhaps driving factors in selecting the platform in the first place). This iterative process and its relationship to experimental design are depicted in Figure 4.1, where the design of the experiment emerges as a result of the research question and the constraints posed by webcam eye-tracking, including spatial and temporal limitations. However, the design is continuously refined as the experimenter explores and tests the available options provided by the system. The objective is to minimize the effects of the method's limitations on the study. Unlike experiments conducted in controlled laboratory environments, our experiences have shown that the relatively unexplored online webcam eye-tracking domain requires extensive pilot testing to identify the optimal set of system options that can effectively address a specific research question.

4.2.2 Opportunities to work around or within the constraints of webcam eye-tracking

In general, webcam eye-tracking systems provide choices for experimenters regarding calibration robustness and offer varying degrees of flexibility. This flexibility in a webcam eye-tracking system can serve as a differentiating factor when selecting a webcam eye-tracking system. Experimenters are typically provided with the flexibility to adjust various aspects of the calibration and recalibration processes, as well as the format of recorded gaze data. For instance, the duration of the calibration process differs across systems. RealEye (Wisiecka et al., 2022), for example, offers a 40-point calibration, while Gorilla's (Anwyl-Irvine et al., 2020) Webgazer (Papoutsaki et al., 2016) implementation allows experimenters to choose between a shorter 5-point calibration or a longer 9-point calibration. Labvanced (Finger et al., 2017), on the other hand, provides even more options, including 15-point (∼40 seconds), 55-point (∼2 minutes), 130-point (∼5 minutes) or 175-point (∼7 minutes) calibrations. While webcam eye-tracking systems may have plug and play capabilities with default settings, gaining a practical understanding of the available options and their effects can assist experimenters in minimizing the spatial and temporal limitations associated with webcam eye-tracking.

To minimize the temporal limitations associated with webcam eye-tracking, which primarily stem from a low sampling rate resulting in an insufficient number of gaze samples, the key approach is to prevent the collection of data with a low sampling frequency from the outset. While certain participant efforts, such as closing background applications on their device, may potentially contribute to a higher sampling rate, overcoming temporal limitations is challenging since most factors are beyond experimenter control. These immutable factors include computing power, available resources, webcam frame rate, and internet quality. Therefore, the most effective strategy to address the issue of low-frequency gaze data collection is to restrict participation in the study to individuals who meet a predefined sampling rate threshold. Ideally, this is done as early as possible in the study to not waste excluded participants' time. Alternatively, in specific cases where the research question and the target population warrant it, conducting a preliminary screening of participants to assess their sampling rate may be appropriate. This approach could be particularly relevant for niche samples such as young children or specific clinical populations as opposed to broad crowdsourced populations. The determination of the appropriate threshold will be explored further in the subsequent section following our examination of methods to mitigate spatial limitations.

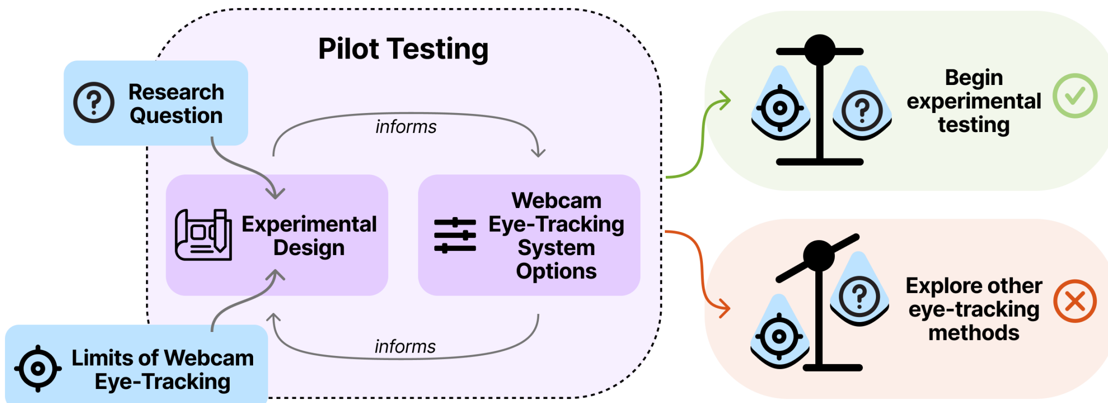


Figure 4.1: A flowchart depicting the process of determining whether you're ready to begin collecting experimental data. Your experimental design should be built with consideration of both your research question, and with an appreciation for the limits imposed by choosing webcam eye-tracking as your eye-tracking method. Multiple cycles of pilot testing are typically required to understand whether your experimental design and your choice of system options are compatible and give you the ability to adequately answer your research question. However, if after iteratively pilot testing to explore your options, it is not possible to balance the limitations of webcam eye-tracking with your ability to answer your research question (via adjusting your experimental design and the system options available to you), other less-constrained forms of eye-tracking should be explored.

In contrast to the preventative approach used for addressing temporal limitations, spatial limitations in webcam eye-tracking, particularly related to spatial accuracy, offer more opportunities for experimenters to take mitigating actions. These limitations are closely tied to the quality of the gaze prediction algorithm and the understanding that the algorithm itself has inherent limitations contributing to spatial inaccuracies. As such, there are avenues to enhance the accuracy of gaze predictions by providing the algorithm with ample data to make more precise predictions. This can be achieved through several means, such as optimizing the robustness of the calibration process (like varying the number of points used in the calibration procedure as discussed earlier). Additionally, employing a robust and frequent recalibration procedure and implementing online data cleaning processes like drift correction or other spatial filters can contribute to improving gaze prediction quality. Furthermore, efforts like setting a threshold for acceptable spatial error after calibration, or incorporating ongoing checks for the stability of the participant's head position (e.g. the "virtual chinrest" feature in Labvanced, or RealEye's "head movement watcher") can enhance the quality of the data used by the gaze prediction algorithm, leading to improved spatial accuracy of the recorded gaze data.

Determining the optimal set of choices among the available options requires careful consideration of the research goal, the spatial and temporal limitations of the system, and the specific experimental design. As depicted in Figure 4.1, this process is iterative and recursive, often involving exploration and testing of various settings to ensure that the research question can be effectively addressed. It's important to recognize that every research question is unique, and the role and purpose of webcam eye-tracking data can vary greatly across different contexts and use cases. Consequently, what may be the most appropriate settings for one study might not be suitable for another. For example, in short studies where eye-tracking is primarily used only to broadly assess participants' engagement (e.g. determining whether they were looking at the screen during the presentation of critical information), opting for minimal calibration and not excluding participants with low sampling rates might be sufficient. Or, in studies involving specific populations where participant recruitment poses challenges, it may be most practical to be more lenient when setting thresholds, acknowledging that data quality issues may need to be addressed in the post-data collection phase. Ultimately, the selection of settings should align with the specific research objectives, taking into account the interplay between spatial and temporal constraints, and the particular demands of the experimental design. By carefully considering these factors and conducting iterative pilot testing, researchers can arrive at the most appropriate settings that enable them to obtain reliable and meaningful webcam eye-tracking data.

In the studies discussed in the During and After Data Collection sections, a set of relatively restrictive settings was applied, including a 5 minute, 130-point calibration, a requirement of less than 7% calibration error, re-calibration every 5 trials, and a minimum system performance of over 10 Hz. In turn, a significant number of crowdsourced participants 'returned' the study (see Figure 4.2), likely due to system performance limitations or dissatisfaction with the complexity of the experiment. However, considering our research goals, we deemed this trade-off to be reasonable in order to obtain meaningful data for our particular use case (as outlined in the After Data Collection section). In the upcoming section, we will elaborate on the iterative pilot testing process that we engaged in to achieve our effective implementation of webcam eye-tracking.

4.2.3 Testing your design with consideration of the constraints of webcam eye-tracking

Crafting an interesting research question and implementing webcam eye-tracking into your experiment is an excellent starting point. However, it is important to recognize that these efforts are akin to hypotheses requiring rigorous testing. Unlike the controlled laboratory setting, where the use of lab-grade eye-trackers is well-established, webcam eye-tracking is in its infancy as a method. Consequently, its application necessitates more guidance and involvement from the experimenter. While likely familiar in practice to any behavioural experimenters, pilot testing for online experimentation (see Johnson et al., 2021 for general online experiment pilot testing guidance), and especially for experimentation using a technically complex method like webcam eye-tracking, is a highly essential component for successful data collection. Treating the pilot process as an iterative, small-scale feasibility study allows online webcam eye-tracking researchers an opportunity to identify and address any potential issues before releasing the study to a broader sample. In practice, this may involve testing the study on a small cohort (around 3 to 5 crowdsourced or local participants) or reducing the number of trials to evaluate the study's mechanics and flow. Providing multiple opportunities for participant feedback during this stage is crucial, as is communicating the purpose of the pilot study to participants. In the Engaging with Participants subsection, we also highlight the importance of clear instructions - something that itself requires pilot testing.

During the pilot testing phase of our studies, we actively solicited participant feedback due to the potential for technical difficulties at any moment. At the end of the study, we presented participants with a long-form text box to gather their insights and address any issues they encountered. Additionally, participants occasionally provided unsolicited feedback through the crowdsourcing inbox feature. This feedback can serve as valuable indicators of the need for improved task instructions, better troubleshooting guidance, or whether the selected eye-tracking system thresholds need to be re-evaluated. Performing a cursory assessment of participant attrition related to eye-tracking system settings and exploring the effects of these settings on the data can be particularly useful for first-time webcam eye-tracking researchers (note: our Filtering participants in practice subsection goes well beyond a cursory assessment). While challenges may arise during any form of human data collection, addressing unforeseen and untested issues is more manageable in a small pilot study before proceeding to a larger-scale study. By incorporating pilot testing and user feedback into the study design and implementation, researchers can enhance data quality, participant engagement, and overall participant (and experimenter!) satisfaction.

4.2.4 Other considerations before collecting data

Appreciation of gaze data as timeseries data

Once a webcam eye-tracking system has been configured most effectively (as determined by thorough pilot testing), it is essential to consider how gaze data is recorded and stored. Drawing on prior experience with collecting and analyzing time series data, such as other biological signals, can provide valuable insight into setting up the necessary "scaffolding" for high-quality data recording. The raw webcam eye-tracking data typically consists of (X,Y) coordinates representing the predicted gaze location over time. However, there are nuanced differences in this raw data output among various webcam eye-tracking applications. These differences include factors like the spatial (coordinate) and temporal frame of the data, as well as whether the gaze signal is sampled and recorded at a consistent, stable rate. We discuss the implications of these factors when processing the data in After Data Collection. In our case with Labvanced-acquired data, the (X,Y) coordinates are recorded relative to a scaled 800 X 450 pixel-unit frame, matching the coordinate frame of the experiment. Each (X,Y) gaze coordinate is time-stamped using the computer's internal clock (e.g. UNIX time) in a non-stationary and dynamic manner, capturing changes in the predicted gaze location as quickly as the system processes them (as opposed to being recorded at a fixed interval of time). To make sense of this gaze data, it is necessary to have spatial and temporal records of relevant on-screen objects and events.

In practice, capturing changes in the coordinate positions of interactive or dynamic objects is crucial in the spatial domain. For example, tasks that involve participants dragging an object across the screen (e.g. our Object Interaction task described in the During Data Collection section) or scenarios where targets appear in randomly assigned locations require additional spatial data to provide context and meaning to the gaze data. When both the gaze and key objects, whether stationary or dynamic, are recorded within the same coordinate frame, it becomes possible to understand how the gaze interacts with the presented stimuli.

In the temporal domain, recording timestamped event flags is akin to recording object positions. While precise timing of stimulus presentation is typically achieved using specialized equipment (like a ViewPixx monitor, for example) in a laboratory setting, online experimentation has limitations in terms of stimulus presentation precision. Timestamping events helps retain some level of temporal accuracy. This means that even if a stimulus is not presented at the exact desired moment on a given trial, the experimenter knows precisely when it was presented. This information adds further context to the gaze data and enables accurate calculation of measures such as the time to first fixation on a presented stimulus. In both studies described in the During Data Collection section, we recorded the time of every frame change and timestamped participant-generated events like response initiation and completion, and interaction with targets. Having the timing information of these events allowed for the analysis of relative segments of data across all participants, despite variations in their absolute timings due to system fluctuations or the self-paced nature of the task. Unless the number of events becomes so large that it overwhelms the system's ability to record eye-tracking data, it is generally preferable to record more events rather than fewer. Pilot testing provides a good opportunity to confirm that all necessary events are accurately recorded in the data output.

Engaging with Participants - Recruitment & Instructions

Recruiting participants for online webcam eye-tracking experiments can be approached in various ways, but using a crowdsourcing platform offers several advantages, including efficient scaling of recruitment and access to a wider range of participants (Aguinis et al., 2021; Johnson et al., 2021). This approach also enables rapid data collection, as noted by several webcam eye-tracking researchers. For instance, Yang and Krajbich (2021) reported that webcam eye-tracking data collection only lasted a couple of days, whereas they estimated it would have taken several weeks in a laboratory setting. In our studies, we used Prolific (www.prolific.co), an academic research-focused crowdsourcing platform, but other platforms like Amazon Mechanical Turk are worth exploring as well. Traditional methods of participant recruitment may not be suitable for webcam eye-tracking experiments depending on the restrictiveness of the selected eye-tracking settings. As discussed in the Filtering participants in practice subsection, our chosen eye-tracking settings were particularly restrictive and resulted in significant participant drop-off at the outset during threshold checks. Recruitment via campus communications or an undergraduate research pool may be especially inefficient when dealing with this uncontrollable form of attrition (however, for specific research questions, given the local sample, some experimenters may choose to use webcam eye-tracking within the lab, where steps can be taken to ensure thresholds are met). If crowdsourcing is not suitable for your research question, a threshold test study can be employed as an initial screening tool in the recruitment process. This can involve a simple test link to determine whether a participant meets the eye-tracking initiation threshold requirements. While this is not a perfect solution, as participants may change devices or their system processing power may fluctuate, it is a way to protect from inefficient recruitment efforts.

Effectively using a crowdsourcing platform like Prolific necessitates clear communication with participants, given the remote nature of the recruitment process. With Prolific, this begins with the study title and description displayed to potential participants. We have found it highly effective to include specific participation requirements directly in the study title (e.g. "[task title] - webcam required") and further emphasizing them in the text of the study description (both at the beginning and later in the text). The study description should outline the tasks participants are expected to perform and provide troubleshooting solutions for any potential issues. In the

Supplementary Materials (see Appendices B and C), we include an example of the recruitment text we used on Prolific. Pilot testing (as previously discussed), provides an opportunity to receive feedback on these initial participant communications and refine them for clarity. For instance, based on pilot participant suggestions, we added a mention of the requirement to remain still for the virtual chinrest feature to the study description. Another useful tool when running more involved remote experiments is the messaging inbox offered by most crowdsourcing platforms. This allows participants to directly communicate with experimenters if any issues arise. Similar to in-laboratory experimentation, it is important to be easily accessible to participants and address any questions they may have, as clear and responsive communication is crucial for successful data collection.

Clear communication for remote webcam eye-tracking studies extends to providing participants with instructions that are easy to understand. It should be assumed that the average participant is not familiar with eye-tracking, either technically or as a scientific method. Clearly informing participants about the eye-tracking process can encourage their engagement, build trust, and address any potential concerns. For example, participants, both in the lab and online, may have valid privacy concerns regarding the collection and use of their data, especially when it extends beyond button presses or survey responses. While informed consent documents should cover data collection and use, it is helpful to reiterate this information before the webcam turns on. Many webcam eye-trackers, including Labvanced, only record gaze position in numeric form without recording or transmitting facial video data. Additionally, the calibration process should be clearly and unambiguously explained, leaving no room for confusion or questions. For first-time webcam eye-tracking researchers, piloting the calibration process with naive participants for feedback can be beneficial. In our studies, we complemented Labvanced's built-in calibration process with a supplementary warning message just prior to calibration (e.g. "calibration will begin once you press the Next button to proceed"), which helped ensure that participants were not confused or unprepared for this critical procedure. In the Supplementary Materials of Chapter 2 and 3 we include a link to our task instructions (see Appendices B and C.

4.3 During Data Collection

Congratulations - you have developed, piloted, and launched your study online! Now what? The data collection phase of a remote webcam eye-tracking experiment is exciting for reasons like the efficiency and scalability of the collection process, but this phase can also introduce frustrations. In this section, we explain the data collection process in the context of the deliberate choices made in the Before Data Collection section, and highlight how these choices can contend with the typical frustrations of remote data collection. We then give real-world context to these ideas, using past remote webcam eye-tracking experiments (n = 151) to trace the flow of participants through the filtering process.

4.3.1 Filtering participants in theory

Data collection during webcam eye-tracking studies is complicated by two key factors: A) the use of a technically-complex research tool like webcam eye-tracking and B) the online, remote setting, where experimenters have limited control over participants' engagement and commitment (Buchanan & Scofield, 2018; Cheung et al., 2017; Clifford & Jerit, 2014; Johnson et al., 2021).

Some of the challenges associated with online data collection and crowdsourcing platforms more generally include sampling concerns (A. Newman et al., 2021), participant inattention or lack of engagement (Johnson et al., 2021), and susceptibility to fraudulent data and bots (Dupuis, 2019; A. Newman et al., 2021). These have been discussed elsewhere (e.g. Aguinis et al., 2021; Gagné and Franzen, 2023; A. Newman et al., 2021; Sauter et al., 2020; Thomas and Clifford, 2017), and we highly recommend first-time online researchers familiarize themselves with these and other challenges specific to the online context, such as achieving representative sampling on crowdsourced platforms (N. Stewart et al., 2017) and the utility of attention checks (when applicable).

The above works also highlight many benefits to online research and the use of crowdsourcing platforms, including the efficient and cost-effective nature of data collection (Semmelmann et al., 2017), the limited oversight required by the experimenter (M. D. Buhrmester et al., 2018; Peer et al., 2017), and the ease of collecting large sample sizes (Aguinis et al., 2021; Gagné & Franzen, 2023). The accessibility of participants and rapid rate of data collection (e.g. sometimes only hours rather than weeks, Yang and Krajbich, 2021) are excellent in contributing to a large pool of potential participants eager to access studies. However, in reality, the use of a complex method like webcam eye-tracking introduces multiple opportunities for issues that could threaten the quality of data collected, and in turn, the ability to answer the research question. Taken together, these reasons are why the process of obtaining complete raw datasets can be likened to a filtration system, as depicted in Figure 4.2.

Throughout the data collection process, would-be participants are filtered out for a variety of reasons. Determining the parameters of these filters was discussed in the Before Data Collection section. Ideally, these choices have been tested through pilot testing prior to data collection, with their use being driven by the data quality they will afford and that the research question necessitates. For some, this means that a substantial number of would-be participants will be filtered out as was the case in our own studies.

As webcam eye-tracking researchers Yang and Krajbich (2021) have pointed out, "what matters is the final number of subjects, rather than the fraction of recruited subjects". Yet, there *is* a point where the fraction becomes too small, making webcam eye-tracking inefficient and economically irresponsible. Therefore, it's important to appreciate the filtration process we face as webcam eye-tracking researchers, and explore opportunities to minimize the effects of filtration (e.g. costs, time) throughout the data collection process. We present our data collection process as a means of highlighting various approaches that can be taken to filtering participants. The specific processes themselves may be of less relevance and will change across platforms and experimental designs, but our general approaches to filtering are broadly relevant.

Initially, the study is distributed through crowdsourcing platforms like Prolific or M-Turk, or through alternative channels such as undergraduate research pools or email lists. Here, the first round of participant filtering occurs. Despite interested participants having the necessary means to access the study link (i.e. a device with internet access), not all those interested will meet the temporal threshold due to insufficient computing resources. Others lack the necessary hardware (i.e. computer and webcam), and some interested participants fail to proceed from the crowdsourcing platform to the study (linked in the study description on the platform). This initial check results in a significant filtering out of participants, as we quantify later. Therefore, it is crucial to assess participant eligibility for webcam eye-tracking as early as possible to avoid wasting participants' time. For example, platforms like Labvanced perform these checks immediately after the study link has been clicked. Since the temporal resolution threshold is set by the experimenter, it is important to make well-informed choices, as they directly impact the likelihood of participants passing the first filter of eye-tracking initialization (as discussed earlier in the Before Data Collection section). Further, this first layer of filtering, if handled at the very outset of the experiment, and paired with clearly communicated participant requirements, can have important cost implications, minimizing the amount of money spent on unusable or poor quality data.

Once the participant successfully initializes the study, including meeting the temporal requirements of the eye-tracking system, they must pass the second filter, which involves an eye-tracking data quality check for spatial accuracy. This check evaluates how accurately the gaze prediction algorithm estimated the gaze location during the calibration phase, where it is assumed that the true location of the gaze aligns with the known location of the calibration points used (because participants are instructed to look at those points). The calibration period is crucial for calibrating the prediction algorithm to the participant's eyes and environmental factors like ambient lighting. However, it also serves as an opportunity to assess the estimated calibration error and identify potential factors that may hinder accurate gaze predictions, such as reflections on a participant's glasses, for example. The prediction algorithm will always attempt to provide a prediction if given data, even if that data is spurious. In the glasses case, if the screen's reflection inhibits the ability to correctly locate the participant's eyes, the predicted gaze data will fail to represent the gaze in any meaningful way (and will only be a source of noise during data analysis). To address this, experimenters can set a maximum allowable calibration error threshold and decide whether to allow the calibration to be repeated if the threshold is exceeded (based on the research question and other considerations discussed in the Before Data Collection section).

As shown in Figure 4.2, the second filtering stage acts as the boundary between complete and incomplete datasets. It also helps address common challenges of online research, such as participant abandonment, lack of personal responsibility, failed attention checks, or difficulties in understanding or performing the tasks. Employing strategies discussed in the Before Data Collection section, such as interactive instructions and clear communication about study requirements, can help reduce this form of participant attrition. For instance, we explicitly state the requirement to remain still throughout the experiment in our study description on Prolific to minimize surprises and frustration for participants. However, besides these challenges, there are also uncontrollable contributors to participant attrition, inherent to the complexity and remote nature of webcam eye-tracking studies. These include technical difficulties and "frozen" errors that prevent participants from completing the study. In such cases, it is important to be available to troubleshoot issues with participants, even though some problems may be beyond our control, resulting in participants being

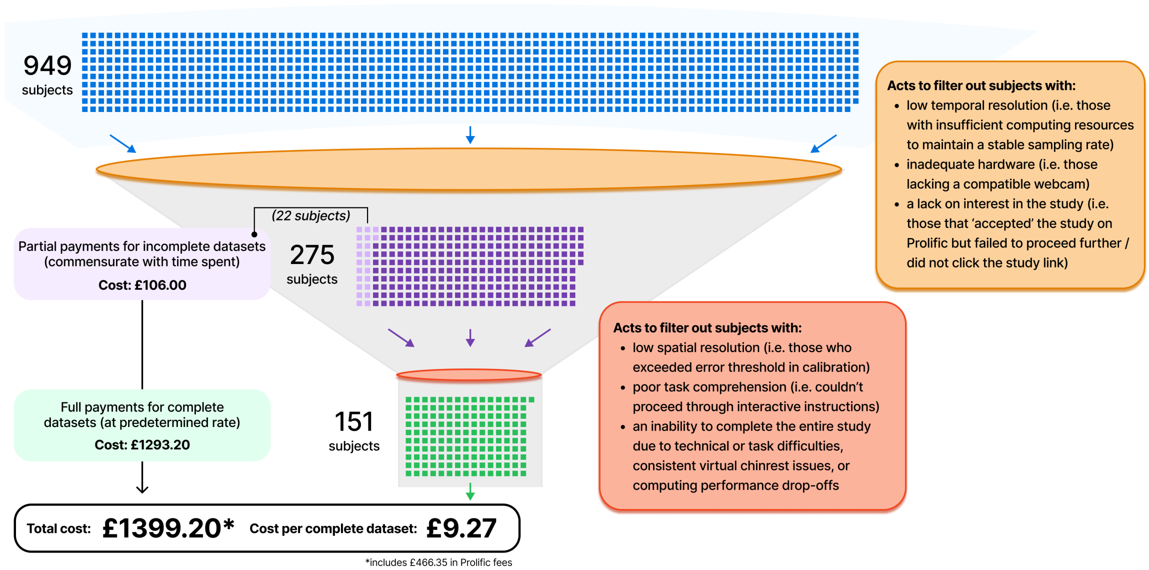


Figure 4.2: A visual depiction of the participant filtering process in practice, showing the outcome of the data collection process of both example studies combined (i.e. Bertrand and Chapman, 2023; Bertrand et al., 2023). Each small square represents one subject at each stage of the filtering process. Explanations of each stage of filtering are provided for additional context, as well as the monetary implications.

advised to end the study without further attempts.

During this second filtering process, participant attrition demands particular attention, especially considering the investment of time made by the participants. Ethical research practices necessitate compensating participants for their time. Minimizing attrition during this stage can therefore have a tangible impact on reducing expenses associated with incomplete datasets that are unlikely to be included in the analysis (depending on the research question). While incomplete datasets due to technical difficulties and other factors are not uncommon even in lab settings, the challenges are amplified in the context of remote webcam eye-tracking studies, as we illustrate further in the next section.

Finally, at the end of the filtering process, we obtain a collection of datasets that represent participants who have successfully completed the study, providing usable eye-tracking data for analysis. Next, we'll discuss how this plays out during real, previously-collected experiments involving remote webcam eye-tracking.

4.3.2 Filtering participants in practice

To provide a practical understanding of the filtration process and add context to the data discussed in the next section, we examine two previously-published large-scale webcam eye-tracking studies (Bertrand & Chapman, 2023; Bertrand et al., 2023). These studies were conducted using the Labvanced online experimentation platform, and participants were recruited with the crowdsourcing platform Prolific. The data collection periods of each study happened within one year of each other, and each study required participants to spend around one hour completing the tasks.

The first study, referred to as the Object-Interaction study (Bertrand & Chapman, 2023), involved dragging and dropping a circle to target locations. The second study, the 2AFC study (Bertrand et al., 2023) focused on decision-making using a cursor. As the authors, we planned to collect 50 complete participants for the Object-Interaction study and 100 complete participants for the 2AFC study. Detailed methods can be found in the respective study publications. Both studies aimed to investigate the coordinated patterns of eye movements and cursor actions, which guided our choice of moderately restrictive thresholds for webcam eye-tracking. For example, we set our initial temporal resolution threshold to a 10 Hz minimum face sampling rate (note: Labvanced uses the face sampling rate as a temporal resolution threshold rather than gaze sampling, as it is an indicator of the system's resources to track the gaze's origin - a component critical for accurate gaze prediction). From here, we will refer to both studies as a combined dataset for the purposes of this discussion.

During the data collection phase, a total of 949 subjects "accepted" the study on Prolific (depicted in Figure 4.2). However, only 275 (29%) subjects actually started the study, making it past the first filter. A breakdown of which filter features contributed the most to this drop-off of 674 (71%) subjects cannot be determined due to the lack of data from Prolific and Labvanced. That is, Labvanced does not store any information if a subject fails to pass the eye-tracking initialization (which encompasses the hardware check and temporal resolution check), making it impossible to determine whether participants failed to meet the requirements or failed to click the study link. It might be reasonable to assume that most crowd-sourced participants intended to participate, and that the drop-off was therefore likely due to the insufficient processing power of the majority of interested Prolific participants. However, this kind of study is also unlike most studies performed on the crowdsourcing platform. If participants employ a strategy of accepting any limited time study offerings before assessing their eligibility (which appeared to be the case based on the almost instant speed that participants would 'accept' the study once it was live on the platform), this amount of participant filtering is likely not only attributable to insufficient processing power. Further, although we applied an additional participant filter using Prolific's option to require a webcam device, this screening criteria is assessed only when participants self-select that they have a webcam device when initially registering on the platform, rather than being some kind of enforced live check prior to the participant accepting the study. Therefore, participants may only realize their lack of appropriate hardware after accepting the study invitation.

Given the substantial number of participants filtered out of the study at this early stage, we underscore the importance of clear and effective communication with participants throughout the entire research process. This is particularly vital for essential aspects like the eye-tracking initialization check, where, in our case, the only recourse for unresolved issues was to request participants "return" the study. When employing relatively stringent thresholds to obtain higher-quality webcam eye-tracking data, effective communication becomes even more important. In our case, through pilot testing, we anticipated this level of participant attrition, and recognized that it was likely due to factors beyond participants' direct control, such as their computer processing power. To prevent wasting participants' time and the experimenter's economic and human resources in addressing issues beyond their control, we repeatedly warned participants from the beginning that the experiment might fail to run on their computer, leading to their inability to participate. By conducting this check at the outset and providing multiple warnings, we significantly minimized the time invested by those that couldn't pass the initial filter. Participants who failed the initial eye-tracking initialization could determine their eligibility within 10-15 seconds of launching the study. Consequently, subjects who failed to pass the first filter were not compensated.

Out of the 275 participants who progressed beyond the initial filter, only 151 of them successfully completed the entire study. The remaining 124 participants accessed the study but did not complete it, potentially due to eye-tracking issues or other factors. Specific eye-tracking issues may have included insufficient spatial accuracy of the gaze prediction, as determined by the calibration procedure. In our case, we used a maximum allowable calibration error cut-off of 7% of the participant's screen size. Furthermore, participants may not have completed the study if they experienced recurring difficulties with the virtual chinrest feature or encountered technical problems related to connectivity or processing instability. Participants who passed the first filter but failed to progress beyond the second filter were compensated for their time, based on the timestamped data in the Labvanced records.

In total, we spent £1399.20 in crowdsourcing costs to collect 151 participant datasets during an hour-long webcam eye-tracking experiment. One third of all costs were service fees to Prolific, with the rest of the funds paid to participants for their time. Of the £1399.20, £1293.20 was spent on complete datasets (n=151), and £106.00 was spent on incomplete datasets (only paid to 22 of 124 incomplete participants given their time investment), amounting to $\sim$ 8% in additional costs for unusable data. With consideration of these additional costs, the cost per complete dataset was £9.27. Compared to experiments run in our lab, where we pay participants at a rate of about £6.00 ($10 CAD) per hour, the costs for our crowdsourced webcam eye-tracking data are substantially higher. These differences come almost entirely from the 33.33% service fees to Prolific, which, when excluded amount to a cost of £6.17 per one-hour dataset collected. When considering the difference in experimenter time to collect 151 eye-tracking datasets in the lab (in a serial fashion) as compared to being collected in a parallel fashion with crowdsourcing, these additional costs seem reasonable, and likely are saving lab personnel costs. We do note that these costs represent those only from our data collection phase, and aren't a representation of the costs associated with the piloting phase. As an example of pilot costs, across our two studies, it took 19 pilot subjects to start collecting experiment-quality data. Because the Object Interaction task hadn't yet been designed for online deployment, and was our first online eye-tracking study, all 19 of these pilot subjects were run as part of pilot testing for that task. Conversely, the 2AFC task was built off of a previously-deployed online study, and having understood the implications of the system settings from the Object Interaction task, in the end, although small tweaks were made to the instructions, the pilot process was successful to the point of producing experiment-quality data (and were thus included in the analysis). It is hopefully clear that the costs of piloting are not inconsequential, but also that there are savings to be gained from the continued use and familiarity with running webcam eye-tracking experiments.

Ultimately, our journey from 949 participants to 275, and then further filtered to 151 participants, enabled us to obtain the necessary raw data to address our research questions. The quality and analysis of these remaining 151 datasets will be explored and assessed in the subsequent After Data Collection section.

4.4 After Data Collection

After much effort to design an effective study and carry out the data collection process, we end up with a collection of raw datasets. In our case, at the end of the participant filtering process, we are left with 151 complete raw datasets. In this section, we discuss everything that comes after the raw data collection process. This includes assessing the raw data quality and then processing the data to minimize quality issues and maximize the data's utility in answering the research question at hand.

4.4.1 Assessing raw webcam eye-tracking data quality

First, we assess the quality of our raw data. Beyond its use in this document as an informative look into the quality of data acquired in a true experimental context, this is not unlike those informal assessments we performed during and after the collection of this data (and pilot data). As we've repeated, there are many more unknowns for an experimenter when performing remote webcam eye-tracking than in-lab eye-tracking. Unlike the laboratory, where the quality of the data is generally stable and not something that's often thought about or interrogated, the quality of the data from webcam eye-tracking requires the experimenter's active oversight and assessment. Without appreciating the potential data quality issues in one's raw data, experimenters run the risk of drawing false conclusions, or miss the opportunity to uncover valuable insights. This data quality assessment is also highly informative for guiding the data processing strategy, as discussed later.

We assess the quality of our raw data in two separate dimensions: time and space. The consideration of these dimensions is a theme throughout this document, and continues here. While both components are foundational to eye-tracking data, as discussed, the reader might find their research question particularly aligned with one dimension more than the other. However, the interconnectedness of these two dimensions means that poor quality in one dimension can corrupt the quality of the data for both dimensions. A highly sampled yet spatially insensitive signal is as uninformative as a spatially accurate signal sampled only a few times. Thus, appreciating the quality of the data with respect to both the spatial and temporal aspects is important.

Assessing data in time

First, we examine the temporal characteristics of our raw data. Of primary importance, and always of note in the related literature, is the sheer quantity of gaze samples recorded. Given the lack of experimental control in an online setting and the likelihood of recording noisier data compared to lab conditions, it's important that we have a sufficient number of samples to identify the signal through the noise. Therefore, the sampling rate is perhaps one of the most important metrics in describing the quality of the raw data.

We explore the timing quality of the raw data in more than one way. First, by looking at the average sampling rate of the datasets, and then, by exploring the variability of that sampling rate on two timescales: 1) at the level of variability between consecutive samples of eye-tracking data, and 2) at the level of variability across the duration of the experiment.

How do we assess the average sampling rate? As discussed in the Before Data Collection section, when we move to an online, remote data collection context, we forfeit highly precise instruments, like lab-grade eye-trackers that capture thousands of samples per second. The online nature further challenges our sampling ability as various processes need to be carried out on a limited bandwidth, all operating on the processing power of consumer-grade hardware. In light of these limitations, it is reasonable to adopt a "get as much as you can" approach for eye-tracking data in online studies. Rather than enforcing a fixed sampling rate that could lead to skipped samples or suboptimal performance, we collect as many samples as possible within the defined bounds. This approach is reflected in our use of the 'Effective Sampling Rate' metric, which considers the number of gaze samples collected over programmed eye-tracking recording period.

The effective sampling rate of our raw data could be calculated in more than one way, but we defined this metric as the number of gaze samples collected over the time that we have requested (i.e. implemented) eye-tracking. In our cases, we have programmed eye-tracking to run from the first to last frame of every trial. Therefore, we calculate the effective sampling rate per trial as:

$$\text{Effective Sampling Rate} = \frac{\text{n}_{\text{gaze samples collected}}}{\text{time}_{\text{trial end}} \text{ - } \text{time}_{\text{trial start}}}$$

Across all 151 remote participants (pooled between our two studies) who satisfied the various participation requirements (discussed in the Filtering participants in theory subsection), the average effective sampling rate observed was 21.2 Hz (SE = 0.26, Range = 8.47 to 29.53). Across the literature, this sampling rate for remote webcam eye-tracking of crowdsourced participants is about as expected or moderately higher (reports range from 11.5 Hz in Bánki et al., 2022 to upwards of 40 Hz in Yang and Krajbich, 2021), though various approaches to obtaining sampling rates makes it difficult to make direct comparisons. We stress that our average effective sampling rate metric should not be considered in a vacuum; this is the effective sampling rate achieved when particular constraints (like Labvanced's 'medium-high, 10 Hz minimum face processing' threshold) are used. As outlined in the Before and During Data Collection sections, thresholds chosen will impact the number of people able to participate (illustrated in Figure 4.2), and what we now report is the effective sampling rate of those who made it through the filter (i.e. 151 participants rather than the 949 at the start of the filter). Therefore, the effective sampling rate is somewhat constructed by our design - a function of our choices when building our experiment. Had we been more lenient in our minimum face processing rate requirement, we likely would have increased the number of participants who made it through our filters, but would have decreased the overall effective sampling rate. Conversely, since our chosen system (Labvanced) supported webcam hardware with a maximum sampling rate of 30 Hz, which is typical for consumer webcams, it's likely we could achieve a sampling rate closer to 30 Hz with more restrictive thresholds. In the near future, since 60 Hz webcams are now on the market, it's also likely that webcam eye-tracking systems will evolve to be compatible with these faster cameras as well, supporting even faster webcam eye-tracking sampling rates.

The effective sampling rate is informative in giving a metric to broadly describe the temporal quality of a trial or session of webcam eye-tracking, but we can also ask how consistently the samples are collected both across a given trial or across the entire duration of the study.

The first, and more granular investigation of estimated sampling rate variability comes by exploring the extent to which an estimated sampling rate *actually* samples at a consistent rate. For example does 30 Hz on a given trial actually represents 30 equally spaced out (in time) samples in one second (i.e. every 33 ms), or is it instead composed of samples at all different intervals that, when averaged, represent a 33 ms interval. As a metric, the derived insights here are more nuanced because the eye-tracking system doesn't set or ascribe to a fixed sampling rate. Instead, this measure is included to highlight how much variability could be introduced when operating within all the limitations and confines (and technical complexities) of a crowdsourced, webcam eye-tracking experiment. We use a 2D histogram of every subject's "Estimated Sampling Rate Difference" between their estimated sampling rate and the interval of time between each of their recorded gaze data samples (see Figure 4.3).

While systems will vary and technology will improve, remote online webcam eye-tracking will always contend with the limitations brought about by consumer hardware and browser-based applications. This wide range of variability we see in Figure 4.3, where some subjects have estimated sampling rate differences in +/- 300 ms range, highlights the magnitude of the momentary changes that the eye-tracking sampling experiences, and serves to inform why later-discussed data treatments are worth considering. However, the clustering of data near-zero for all subjects suggest that the majority of the time, the estimated sampling rate is a reasonable representation of the general timing of recorded gaze samples. We also note that in both studies reviewed, cursor-tracking was also enabled and recorded (sampled as fast as possible, but only when a change in cursor position is detected), and this may have effects on the momentary stability of the eye-tracking system (because of time-varied throughput demands) and also the estimated sampling rate decay we see over time (increased demands on local cache resources; see Figure 4.4).

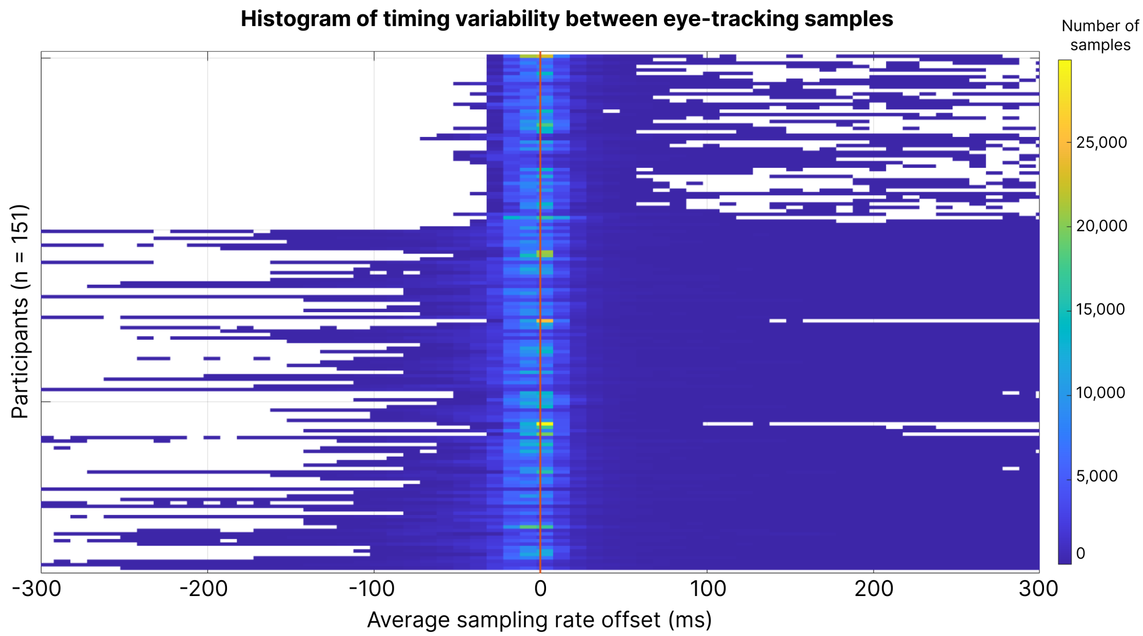


Figure 4.3: Histogram of timing variability between eye-tracking samples, where the variability is calculated as the difference between average sampling rate time difference and the true time differences, in milliseconds. Each participant (n=151) is shown per row, where the Object Interaction cohort (n=51) is shown in the top 51 rows. The colorbar indicates the histogram count of the number of samples for given offsets from the estimated sampling rate.

In our second approach to describing the estimated sampling rate variability, we map out the effective sampling rate calculated per trial, across the duration of the study (Figure 4.4). With our combined datasets having similar average experimental durations (Object-Interaction: M = 45min 07sec, SD = 09min 46sec; 2AFC: M = 1hr 03min 27sec, SD = 10min 09sec) we show all 151 participants' effective sampling rates over the course of their eye-tracking data collection, in 10 minute intervals. A general trend emerges that shows the slowing of the effective sampling rate over the course of the ∼1 hour experiment. This reduced sampling rate may be the result of various factors like the system's backend engineering and handling of data, and/or the ability of the participants' hardware (e.g. processor's temperature regulation) to sustain prolonged, substantial resource use. In the case of a system like Labvanced's, where the processing involved with eye-tracking is handled client-side (i.e. no video data is transmitted from the participant's computer, but is instead processed using their system's resources to generate coordinate gaze data), this continued resource demand likely weighs on a typical participant's hardware. The subject's local cache is also used during the experiment execution and thus would further weigh on performance over time. At least for our platform of choice, webcam eye-tracking may introduce some limitations when trying to record longer studies that exceed 30 or 45 minutes (although this concern will likely diminish rapidly as platform and consumer technology improves). We explore how this diminishing sampling rate may or may not affect data quality further below (like in the magnitude of recalibration error across time in Figure 4.5, for example).

Assessing data in space

While still only considering the data in isolation from the experimental context it was conducted for, following the implementation suggestions in the Before Data Collection section, we can now use the recorded calibration and recalibration error as an indicator of general spatial accuracy. Prior to any experimental trials, all 151

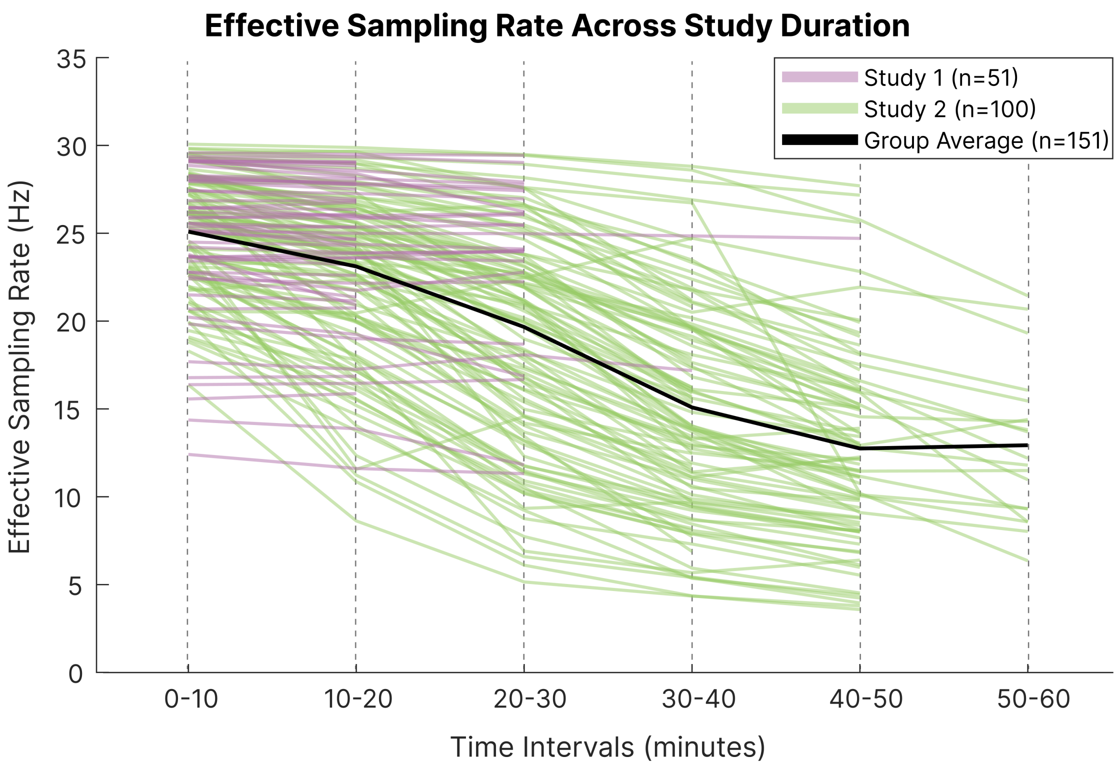


Figure 4.4: The effective sampling rate, in Hz, across 10 minute intervals of the recording duration is shown, with each line representing a single subject (color-coded for experimental cohort) and the group average indicated with a black line.

subjects performed a 5 minute calibration procedure where they fixated targets on the screen from a range of head positions. At the end of this preliminary calibration, a calibration error was calculated that represents the magnitude of error between the algorithmically-predicted gaze location and the known target locations (taken as ground truth). Of course, in practice, this measure, and the taking of the calibration targets as ground truth, is muddied by the limited experimental control of a remote study, but our use of a restrictive calibration error limit (see the Before Data Collection section) provides some reassurance that participants are meaningfully engaging in the calibration and that the subject's environment is suitable for effective webcam eye-tracking (i.e. appropriate lighting and background, limited visual distractions). Figure 4.5A depicts the initial calibration error recorded for each subject within each experimental cohort. This calibration error is an aggregated measure of the magnitude of error of all the calibration points, for X and Y domains separately, measured within the 800 x 450 common pixel unit coordinate frame used by Labvanced. Our use of calibration error as a metric is limited by the error being recorded in an unsigned way (i.e. for 40 pixels of error in the X domain, for example, it's unclear whether that error represents a prediction bias to the left or right). However, Figure 4.5A does expose a bias in our data using a fitted least squares reference line, where the initial calibration shows an unbalanced error contribution of more error in the X domain than the Y domain. Given the general properties of computer screens being wider than they are tall, it is not necessarily surprising that we see a least squares line that skews to more error in the X domain than Y, but we share this plot more for information than to generate definitive conclusions. It's worth noting that our imposed restrictions on calibration error (7% in our case) when operating in the Labvanced frame (800 x 450 pixel units) translates to 56 pixels in screen width and 31.5 pixels in height, however the 7% threshold is calculated on the diagonal distance (Euclidean) and thus explains why we see data in Figure 4.5 A that has a magnitude larger than 56 pixels in X or 31.5 pixels in Y but not exceeding either of these in both components.

Our general trend of X>Y gaze prediction error further carries over into the recalibration findings. Here, in Figure 4.5B, we also plot the complete cohort of data (n=151), separated by study. Importantly, the number of recalibrations performed varies greatly between the two studies (recalibration occurs every 5 trials in both studies but subjects perform ∼5 times more trials of a much shorter duration in the 2AFC Task). All data shown are averaged recalibration values. Interestingly, performing only 10 vs 56 recalibrations appears to have little impact on the magnitude of the recalibration error, or the relationship between X and Y error. We further employ a 10 minute interval sampling like the sampling rate results to explore whether the recalibration error changes over time. This average recalibration error is measured in Euclidean pixel units, taking subject averages of the Euclidean distance of each calibration period over 10 minute intervals. From Figure 4.5C, we see an upward trend to the average magnitude of recalibration error over the duration of the experiment, suggesting that the spatial accuracy may decay over time when performing hour long experiments. However, this result may also be a consequence of a broader data quality decrement when remotely running longer experiments where prolonged attention is required.

4.4.2 Data Processing

All data presented thus far has been wholly raw data. In practice, despite our best efforts to create a sound experiment and run an efficient data collection process, the quality of the eye-tracking data collected can only be controlled to a certain extent. We therefore rely on exercising more experimental control post data collection, where, when appropriate, we use more liberal trial and subject rejection thresholds than we might employ in the lab. Given the vast span of research questions that can be answered with webcam eye-tracking, this guide pertains primarily on how to run an online webcam eye-tracking experiment, but we share some general data processing

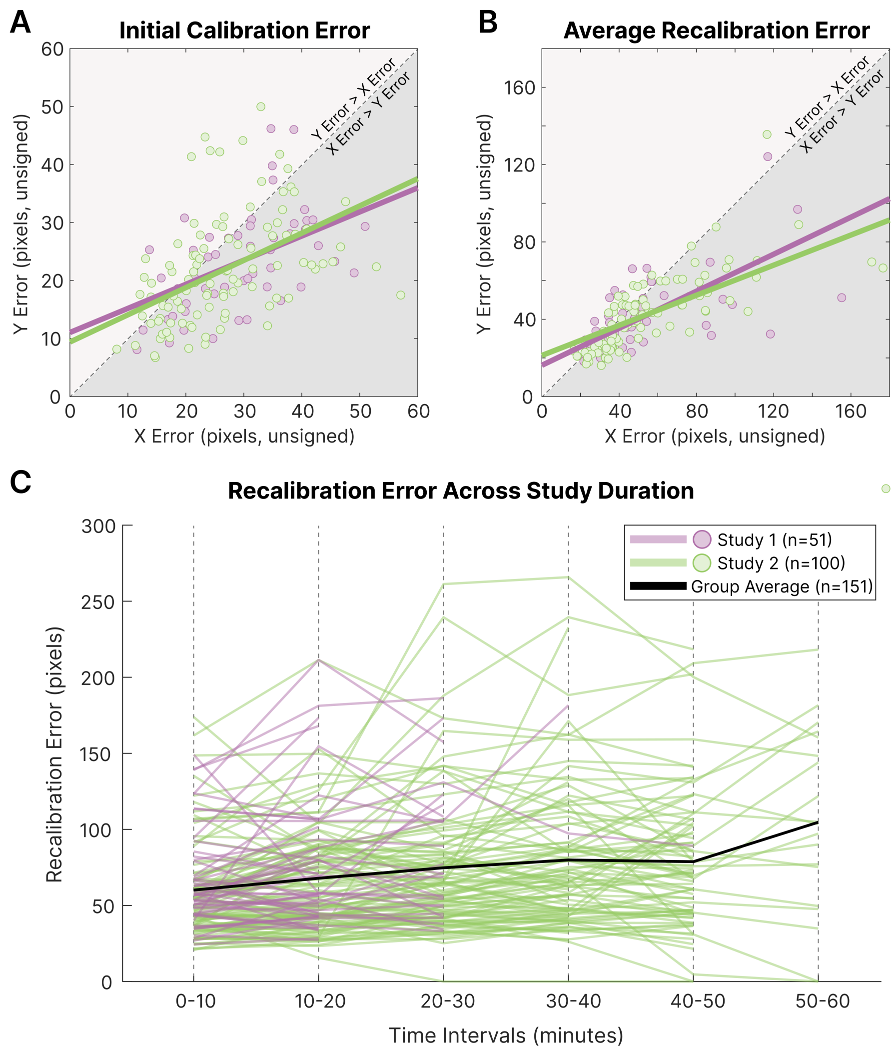


Figure 4.5: Various measures of webcam eye-tracking spatial error from the two experiments. A) The initial calibration error is shown, with individual calibration errors indicated by circles, color-coded with the experimental cohort. Initial calibration error is determined after the full calibration procedure is complete, prior to any experimental recording. Error is in unsigned pixel space, using a common coordinate frame of 800 x 450 pixels. A diagonal line separates the two halves of the plot to highlight the areas where there's more error in the X domain or in the Y domain. Two fitted least squares reference lines are shown, per cohort. B) The average recalibration error is shown, where the unsigned error in both the X and Y domain has been averaged over all the recalibrations performed throughout the experiment. All other aspects of the plot are identical to plot A. C) Recalibration error across the study duration is shown, where the average recalibration error is computed for each 10 minute interval of recording time. Error here is calculated by combining the error in the X and Y domains with a euclidean distance measure. Each line represents an individual's average euclidean distance error across the time intervals, colour-coded for cohort. The black line indicates the grand average of the entire n = 151 group.

ideas and practical approaches we have found useful over our online experimentation journey.

We estimated the sampling rate from eye-tracking data that was collected without a fixed sampling rate (see the During Data Collection section and Figures 4.3 and 4.4). A simple processing step to use from the outset is to upsample (or re-sample) your eye-tracking data to a common sampling rate for all the collected data. For our purposes, because we also record cursor position (which is recorded at a much higher sampling rate than eye-tracking data for our datasets), and use the cursor + eye data together in joint measures, we choose a 60 Hz common sampling rate, and linearly interpolate our eye data in the missing frames. A linear interpolation is an oversimplification of the velocity profile of the saccade, but, at present and as previously discussed, research on saccade dynamics is beyond the scope of current webcam eye-tracking capabilities. Given the variability within and between subjects' estimated sampling rates (see Figure 4.3), it would be best to employ an approach that would protect against data with too much time between sequential samples, or not enough samples in general. Depending on your study design and research question, this sampling threshold will vary, but consider, for example, how uninformative only three data points over a five second span would be. There will be almost entirely interpolated data in the processed signal, and this could easily mislead or confuse the data story. This upsampling approach is like an inverse to a filtering approach which most lab-based eye-tracking scientists would use to smooth out their data. In the upsampling case for webcam eye-tracking data, we 'smooth' our webcam eye-tracking data by filling in gaps instead of dampening noisy outliers in 1000 Hz lab eye data.

Besides pertaining to raw data, until now, all measures presented have also been ones without consideration of the experimental tasks being performed by the participants. Of course, for our purposes, generic measures will afford greater general applicability for our audience's various types of experimental designs, but in practice, task-specific data quality assessments are typically the more interesting investigations when taking stock of your raw data output. In general, a guiding principle of exercising a healthy amount of skepticism about the data is one worth following. Given all the limitations of webcam eye-tracking discussed at length throughout, becoming familiar and comfortable with your eye-tracking data in the form of visual inspection (plots, heatmaps etc.) or extracting metrics for practical sanity checks is important. For example, Figure 4.6 is a reproduction from Bertrand and Chapman (2023) where we performed a basic validation of our data quality and data processing approach by measuring the amount of time spent fixating at each of our 4 areas of interest when that area was an active (relevant) or inactive (non-relevant) area for interaction. We used the expected pattern (significantly more fixation at relevant vs irrelevant sites) as a basic proof that later results were grounded in legitimate and valid eye-tracking data.

When beginning to assess quality, consider what values (or patterns of values) you might extract that would give you confidence to proceed with the analysis procedure (or, alternatively, indicate the need for creative data-driven cleaning treatments). Simple measures like time to first fixation to a key target or fixation duration can be useful here, and when combined with other task-agnostic measures like calibration error, may inform first-pass analysis choices like what constitutes a reasonable bounding area about an AOI for metric generation purposes.

Another piece of guidance for data processing, but more as a principle, is to consider in what ways you might be able to discern completely random or erroneous data for cases where the participant may have left the view of the webcam, or are focusing their gaze elsewhere. While there are various ways to avoid this happening during the data collection by implementing things like a virtual chinrest, or designing the experiment to not advance without user input etc., it's important to consider how you might be able to use features within the data to identify trials (or subjects) not suitable for analysis. Some of these ideas were mentioned above when referencing becoming familiar with the data and performing sanity checks; these types of sanity

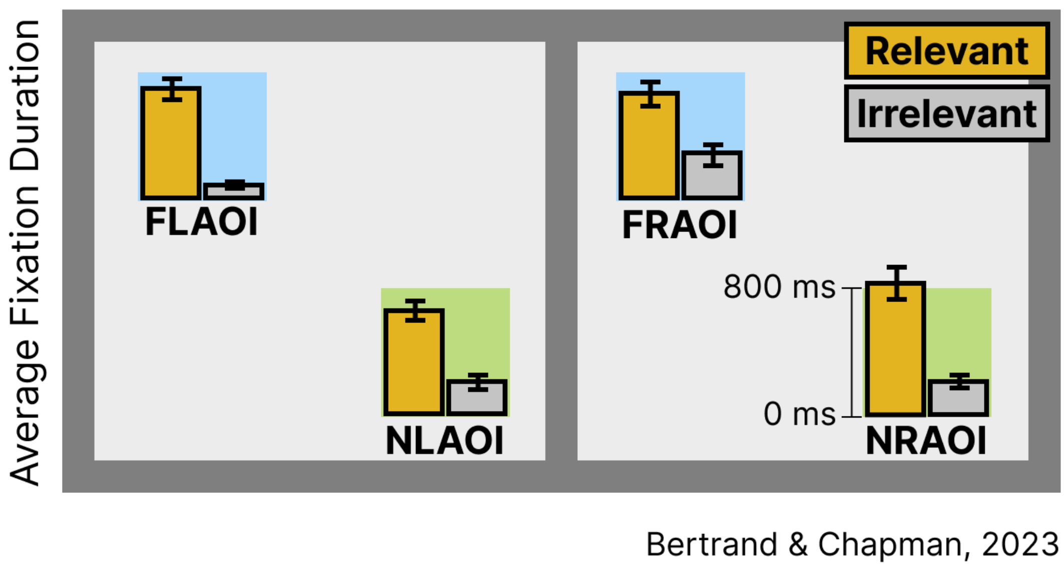


Figure 4.6: An example from the object interaction task (Bertrand & Chapman, 2023) of how task-specific measures can be generated as a means of validating the quality of your collected data. In this case, we measured the average fixation duration at each of the four target AOIs (Far Left: FLAOI, Near Left: NLAOI, Far Right: FRAOI, Near Right: NRAOI) across the task's eight movements when that AOI was relevant (yellow bar; a pick-up or drop-off site) or irrelevant (grey bar; not a pick-up or drop-off site) to the movement. Our results indicated significantly longer fixations on relevant AOIs than irrelevant AOIs, following the proven and established pattern of only looking at task-relevant areas. A reproduction from Bertrand and Chapman (2023)

checks are also reasonable to employ as rejection criteria. In our example above where we classify areas as task relevant or irrelevant, we use this classification to also identify trials and/or subjects that violate this pattern entirely - indicating faulty or unreliable gaze predictions that shouldn't be considered in analyses. In our case, this resulted in analyzing only 29 of the 51 datasets for the Object Interaction study. Determining the ideal criteria requires an appreciation for your research question and your experimental design, and visually inspecting the data can help to clarify a reasonable approach (i.e. 'bad' or strange coordinate gaze data is not obvious numerically until plotted). This criteria is also sometimes important to ensure you have sufficient data to answer your research question. For example, in a task like our 2AFC task, because we wanted to analyze gaze as it related to the two presented choices, we decided to exclude trials where the gaze was never recorded as falling on or around the two choice areas (even when we considered a wider, adjusted area bound). Striking the right balance is important to not reject 'good' data while not effectively accepting noise or 'bad' data. For this particular experiment, this meant we analyzed 97 of 100 participant datasets. Clearly, different research questions will inform these rejection procedures - our two studies varied significantly in the rejection rate of datasets for analysis.

What if these preliminary data processing strategies are still presenting data that is difficult to make sense of? There are a number of more complex or advanced approaches to combat various challenges. For example, our object interaction study (n=29; Bertrand and Chapman, 2023) ended up requiring substantial data transformations by using a k-means clustering approach to spatially bin the data into 4 areas of interest because of spatial distortions (full procedure illustrated and described in the Supplementary Materials of Bertrand and Chapman, 2023, see Appendix B). We also supplemented our relevant/irrelevant proof (see Figure 4.6) with a more complex spatial probability analysis for further validation of our data transformation processes (also in the Supplementary Materials of Bertrand and Chapman, 2023, see Appendix

B). Application of post-hoc fixation detection algorithms on particularly noisy raw gaze data may also be fruitful in distilling data into less quantity and more quality gaze representations. Linearly transforming gaze data based on known in-task anchor locations that can be confidently and reliably used (because of task designs that may require careful monitoring at that location, for example) can also be employed as a pseudo post-hoc drift correction. Though it may require more creative and scrappy strategies than in the lab, if the experiment has been designed and implemented effectively, there will generally be some approach to overcome data quality hurdles that will still allow for powerful answers to your research question. As technology continues to advance, these data quality challenges will become less pronounced, and solutions to the challenges will no doubt improve too.

4.5 Summary of Recommendations

Taken together, we hope this guide will assist in developing and running high-quality webcam eye-tracking experiments. We were motivated to share these ideas after learning through trial and error, effectively performing pilot tests on the method itself. Now, after learning some of the nuances of this methodology, we hope we can share our learnings and knowledge to make webcam eye-tracking more accessible for experimenters to implement. We hope the reader is excited about webcam eye-tracking, and takes away the key messages in Figure 4.7 for each stage of the research process.

4.6 Conclusion

If nothing else is taken away from this paper, we hope the reader sees the importance of engaging in a thorough and iterative pilot testing phase to understand whether it's possible to design and build a webcam eye-tracking experiment that is capable of answering your research question. This document is meant to serve as a practi-

Recommendations for online webcam eye-tracking experiments

Preparation:
- Learn and understand webcam eye-tracking limitations.
- Adjust experimental design for spatial and temporal constraints.

Calibration and Recalibration:
- Explore the available calibration settings.
- Consider the optimal calibration choices for your research question: more time calibrating will likely lead to better data quality.
- Perform recalibrations throughout the experiment, at least once every 5-10 minutes.

Pilot Testing and Recruitment:
- Conduct iterative pilot testing to determine implications of your chosen settings and design.
- Recruit participants efficiently through crowdsourcing platforms.
- Clearly communicate study requirements, webcam eye-tracking process, and task procedures to participants.

Data Collection and Costs:
- Filter out ineligible participants (e.g. lacking hardware) as early as possible to limit collection of incomplete datasets.
- Set calibration error thresholds for data quality.
- Maintain communication and address technical issues.
- Factor in costs for incomplete datasets.

Assess Raw Data Quality:
- Evaluate raw data quality in time and space.
- Examine temporal characteristics and consistency (e.g. effective sampling rate and its variation).
- Assess spatial accuracy through calibration and recalibration errors.

Data Processing and Assessment:
- Consider resampling eye-tracking data for analysis.
- Perform task-informed quality checks and validation.
- Consider rejecting trials violating expected gaze patterns.
- Explore advanced data treatments (e.g. AOI clustering) to address quality issues.

Continuous Improvement:
- Work within constraints through effective design and iteration.
- Consider the costs of learning as investments for future studies.

Figure 4.7: Summary of recommendations for online webcam eye-tracking experiments

cal guide for experimenters, but is limited in scope to the more important pieces of information for performing a successful experiment. Outside of this more practical context, we urge experimenters to also explore further considerations to the method, including implications for participant privacy and security. We also are excited for this eye-tracking method to continue progressing beyond laptop and desktop computer webcams to other consumer cameras like smartphones (actively being explored, e.g. Namnakani et al., 2023). For the most part, extensions beyond webcams to these other types of consumer cameras still present similar challenges for behavioural experimentalists, and our guide will still be relevant for these cases. We also believe there are exciting opportunities for more supportive crowdsourcing environments that enable better participant filtering earlier in time to further limit costs and time. In the end, we're excited for the future of webcam eye-tracking and hope you are able to enjoy the process of contributing to its use in your future webcam eye-tracking research.

4.7 References

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations [Publisher: SAGE Publications Inc]. *Journal of Management*, *47*(4), 823–837.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407.

Bäck, D. (2006). *Neural network gaze tracking using web camera*. Institutionen för medicinsk teknik. Retrieved June 6, 2023, from https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-5579

Bánki, A., de Eccher, M., Falschlehner, L., Hoehl, S., & Markova, G. (2022). Comparing online webcam- and laboratory-based eye-tracking for the assessment of infants' audio-visual synchrony perception. *Frontiers in Psychology*, *12*, 733933.

Bertrand, J. K., & Chapman, C. S. (2023). Dynamics of eye-hand coordination are flexibly preserved in eye-cursor coordination during an online, digital, object interaction task. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–13.

Bertrand, J. K., Zuk, A. A. O., & Chapman, C. S. (2023, June 7). Continuous measures of decision-difficulty captured remotely: II. webcam eye-tracking reveals early decision processing [Pages: 2023.06.06.543799 Section: New Results].

Bogdan, P. C., Dolcos, S., Buetti, S., Lleras, A., & Dolcos, F. (2023). Investigating the suitability of online eye tracking for psychological research: Evidence from comparisons with in-person data using emotion–attention interaction tasks. *Behavior Research Methods*.

Bruno, A., Tliba, M., Kerkouri, M. A., Chetouani, A., Giunta, C. C., & Çöltekin, A. (2023). Detecting colour vision deficiencies via webcam-based eye-tracking: A case study. *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, 1–2.

Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods*, *50*(6), 2586–2596.

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of amazon's mechanical turk, its rapid rise, and its effective use [Publisher: SAGE Publications Inc]. *Perspectives on Psychological Science*, *13*(2), 149–154.

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49–62.

Cheung, J. H., Burns, D. K., Sinclair, R. R., & Sliter, M. (2017). Amazon mechanical turk in organizational psychology: An evaluation and practical recommendations. *Journal of Business and Psychology*, *32*(4), 347–361.

Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? an experimental comparison of data quality in laboratory and online studies [Publisher: Cambridge University Press]. *Journal of Experimental Political Science*, *1*(2), 120–131.

Delabarre, E. B. (1898). A method of recording eye-movements [Publisher: University of Illinois Press]. *The American Journal of Psychology*, *9*(4), 572–574.

Dupuis, M. (2019). Detecting computer-generated random responding in questionnaire-based data: A comparison of seven indices, 10.

Ehinger, B. V., Groß, K., Ibs, I., & König, P. (2019). A new comprehensive eye-tracking test battery concurrently evaluating the pupil labs glasses and the EyeLink 1000 [Publisher: PeerJ Inc.]. *PeerJ*, *7*, e7086.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: A unified JavaScript framework for online studies. *International Conference on Computational Social Science*, *1*(1), 1–3.

Gagné, N., & Franzen, L. (2023). How to run behavioural experiments online: Best practice suggestions for cognitive psychology and neuroscience. *Swiss Psychology Open: the official journal of the Swiss Psychological Society*, *3*(1), 1.

Greenaway, A.-M., Nasuto, S., Ho, A., & Hwang, F. (2021). Is home-based webcam eye-tracking with older adults living with and without alzheimer's disease feasible? *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, 1–3.

Hutt, S., Wong, A., Papoutsaki, A., Baker, R. S., Gold, J. I., & Mills, C. (2023). Webcam-based eye tracking to detect mind wandering and comprehension errors. *Behavior Research Methods*.

Johnson, B. P., Dayan, E., Censor, N., & Cohen, L. G. (2021). Crowdsourcing in cognitive and systems neuroscience [Publisher: SAGE Publications Inc STM]. *The Neuroscientist*, 10738584211017018.

Levine, J. L. (1981). *An eye-controlled computer*. IBM Research Division, TJ Watson Research Center.

Madsen, J., Júlio, S. U., Gucik, P. J., Steinberg, R., & Parra, L. C. (2021). Synchronized eye movements predict test scores in online video education [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *118*(5), e2016980118.

Meng, C., & Zhao, X. (2017). Webcam-based eye movement analysis using CNN [Conference Name: IEEE Access]. *IEEE Access*, *5*, 19581–19587.

Namnakani, O., Abdrabou, Y., Grizou, J., Esteves, A., & Khamis, M. (2023). Comparing dwell time, pursuits and gaze gestures for gaze interaction on handheld mobile devices. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–17.

Newman, A., Bavik, Y. L., Mount, M., & Shao, B. (2021). Data collection via online platforms: Challenges and recommendations for future research. *Applied Psychology*, *70*(3), 1380–1402.

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163.

Robal, T., Zhao, Y., Lofi, C., & Hauff, C. (2018). Webcam-based attention tracking in online learning: A feasibility study. *23rd International Conference on Intelligent User Interfaces*, 189–197.

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences*, *10*(4), 251.

Saxena, S., Lange, E., & Fink, L. (2022). Towards efficient calibration for webcam eye-tracking in online experiments. *2022 Symposium on Eye Tracking Research and Applications*, 1–7.

Schneegans, T., Bachman, M., Huettel, S., & Heekeren, H. (2021). *Exploring the potential of online webcam-based eye tracking in decision-making research and influence factors on data qualit* (Type: article).

Schröter, I., Grillo, N. R., Limpak, M. K., Mestiri, B., Osthold, B., Sebti, F., & Mergenthaler, M. (2021). Webcam eye tracking for monitoring visual attention in hypothetical online shopping tasks [Number: 19 Publisher: Multidisciplinary Digital Publishing Institute]. *Applied Sciences*, *11*(19), 9281.

Semmelmann, K., Hönekopp, A., & Weigelt, S. (2017). Looking tasks online: Utilizing webcams to collect video data from home. *Frontiers in Psychology*, *8*(1582), 1–11.

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451–465.

Sewell, W., & Komogortsev, O. (2010). Real-time eye gaze tracking with an unmodified commodity webcam employing a neural network. *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, 3739–3744.

Shehu, I. S., Wang, Y., Athuman, A. M., & Fu, X. (2021a). Paradigm shift in remote eye gaze tracking research: Highlights on past and recent progress. In K. Arai, S. Kapoor, & R. Bhatia (Eds.), *Proceedings of the future technologies conference (FTC) 2020, volume 1* (pp. 159–189). Springer International Publishing.

Shehu, I. S., Wang, Y., Athuman, A. M., & Fu, X. (2021b). Remote eye gaze tracking research: A comparative evaluation on past and recent progress [Number: 24 Publisher: Multidisciplinary Digital Publishing Institute]. *Electronics*, *10*(24), 3165.

Skovsgaard, H., Agustin, J. S., Johansen, S. A., Hansen, J. P., & Tall, M. (2011). Evaluation of a remote webcam-based eye tracker. *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications*, 1–4.

Slim, M. S., & Hartsuiker, R. J. (2022). Moving visual world experiments online? a web-based replication of dijkgraaf, hartsuiker, and duyck (2017) using PCIbex and WebGazer.js. *Behavior Research Methods*.

Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences*, *21*(10), 736–748.

Stone, S. A., & Chapman, C. S. (2023). Unconscious frustration: Dynamically assessing user experience using eye and mouse tracking. *Proceedings of the ACM on Human-Computer Interaction*, *7*, 168:1–168:17.

Thomas, K. A., & Clifford, S. (2017). Validity and mechanical turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, *77*, 184–197.

Vos, M., Minor, S., & Ramchand, G. C. (2022). Comparing infrared and webcam eye tracking in the visual world paradigm [Accepted: 2023-01-03T13:11:04Z Publisher: University of California Press].

Wisiecka, K., Krejtz, K., Krejtz, I., Sromek, D., Cellary, A., Lewandowska, B., & Duchowski, A. (2022). Comparison of webcam and remote eye tracking. *2022 Symposium on Eye Tracking Research and Applications*, 1–7.

Wong, A. Y., Bryck, R. L., Baker, R. S., Hutt, S., & Mills, C. (2023). Using a webcam based eye-tracker to understand students' thought patterns and reading behaviors in neurodivergent classrooms. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 453–463.

Xu, P., Ehinger, K. A., Zhang, Y., Finkelstein, A., Kulkarni, S. R., & Xiao, J. (2015, May 20). TurkerGaze: Crowdsourcing saliency with webcam based eye tracking. Retrieved August 24, 2022, from http://arxiv.org/abs/1504.06755

Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, *16*(6), 1485–1505. Retrieved August 24, 2022, from https://journal.sjdm.org/21/210525/jdm210525.html

Zhang, X., Sugano, Y., & Bulling, A. (2019). Evaluation of appearance-based methods and implications for gaze-based applications. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13.

Zhao, Y., Lofi, C., & Hauff, C. (2017). Scalable mind-wandering detection for MOOCs: A webcam-based approach. In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, & M. Pérez-Sanagustín (Eds.), *Data driven approaches in digital education* (pp. 330–344). Springer International Publishing.

Zheng, C., & Usagawa, T. (2018). A rapid webcam-based eye tracking method for human computer interaction [ISSN: 2475-7896]. *2018 International Conference on Control, Automation and Information Sciences (ICCAIS)*, 133–136.

Chapter 5

Discussion

The purpose of this thesis was to explore the utility of webcam eye-tracking for human behavioural research. As you've now read, throughout the three projects, we've taken a couple different approaches to understanding utility. In Chapters 2 and 3, we chose distinct domains to use as testbeds for webcam eye-tracking, assessing the method's utility in terms of its practical application and ability to sensitively capture key domain-specific behaviours. Our fourth chapter takes a different approach to exploring webcam eye-tracking's utility, offering instead a contribution to its utility by sharing practical knowledge for employing the method in online behavioural research, and by providing evidence-based recommendations that make the method more accessible for researchers.

In the following sections, I provide an overview and contextualization of the studies conducted within this thesis. Chapters 2 and 3 are presented in an order that first summarizes the extent to which webcam eye-tracking's utility was established based on the key domain-specific indicators I presented in the Introduction (Chapter 1). Then, following these summaries, the utility discussion is bolstered by revisiting the additional empirical contributions the studies made (that webcam eye-tracking afforded us). We further discuss the studies' contributions by situating our results in the existing literature. We conclude the discussions of Chapters 2 and 3 by providing an honest reflection about the process of determining if and how it might be possible to derive meaning from the raw webcam eye-tracking data. With Chapter 4 being a methodological contribution built upon the first two experiments, I briefly summarize its contribution and situate its utility in the current literature. The rest of the discussion shifts towards an exploration of utility as a concept, paired with an exploration of the future of webcam eye-tracking through both optimistic and pessimistic lenses. I close out this thesis with a personal reflection on my journey and experiences with webcam eye-tracking as a method.

5.1 Study Summaries

5.1.1 Chapter 2 - Assessing webcam eye-tracking utility in digital object interactions

Chapter 2 endeavoured to understand the utility of webcam eye-tracking as a tool to explore the persistence of eye-hand coordination patterns within digitized object interactions. Specifically, we wanted to probe whether webcam eye-tracking could identify the familiar 500 ms eye-leading-hand (here, eye-leading-cursor) pattern, commonly observed in real-world object interactions (e.g. Lavoie et al., 2018). Operating within a fully remote experimental setup, we successfully deployed a digital drag-and-drop cursor movement task with webcam eye-tracking, gathering complete datasets from 51 participants who used their own web-cameras as eye-trackers. As we presented in Chapter 2 and discussed later, substantial data processing efforts were made to enable a thorough exploration of the gaze data. These efforts proved fruitful, making it possible to not only explore the gaze data, but also reveal the unique and robust 500 ms eye-leading cursor pattern across all eight movements of our digital interaction task.

The significance of Chapter 2 extends beyond a demonstration of webcam eye-tracking's utility in capturing rich visuomotor behaviours from online participants. It also afforded us the opportunity to unveil novel empirical insights about the adaptability of the visuomotor system in a digital domain. When we quantified eye-cursor latencies during object pick-up and drop-off events, we found that the real-world's 500 ms "look ahead" behaviour manifested differently in digital interactions. Preceding the cursor's object pickup, the eyes mirrored real-world behaviour, arriving at the object approximately 500 ms prior to the cursor. However, due to the rapid nature of dragging a digital object on-screen compared to physical manipulation, the limited time span between the pick-up and drop-off precluded the eyes from having their requisite 500 ms of anticipatory time at the drop-off site. Instead, the eyes arrived at the drop-off location only ∼100 ms prior to the cursor. Remarkably, even after the drop-off, the gaze remained fixated on the drop-off site for an additional 400-500 ms. This surprising finding, revealed with webcam eye-tracking, hints at the existence of a fundamental ∼500 ms window for visual information processing during interactions, both digital and real-world, with the visuomotor system displaying adaptive flexibility in accelerated digital interactions.

These results therefore align with and extend existing literature from both the real-world and digital domains. Returning back to the real-world domain that inspired this work, as has been stated, the presence of the 500 ms eye-leading-cursor visuomotor pattern is consistent with Lavoie et al. (2018) and others' eye-leading-hand pattern (Hayhoe, 2000; Land & Hayhoe, 2001; Land et al., 1999). In turn, this resembles patterns that arise from other real-world, visually-guided tasks like walking (Land, 2006; Patla & Vickers, 2003), keyboard typing (Butsch, 1932), and music playing (Furneaux & Land, 1999). Further, the observed general flexibility of the visuomotor system mirrors real-world instances where imposed time (Deconinck et al., 2011), accuracy (Rand & Stelmach, 2010), or kinematic constraints (Johansson et al., 2001; Pelz et al., 2001) elicit adaptative eye-hand coordination patterns. Our most novel empirical finding was uncovering the consistent, prolonged (∼500 ms) gaze fixation after drop-off. In addition to adapting to the rapidity of mouse movements, this extended fixation might also relate to the limited sensory (haptic) feedback of cursors compared to the rich haptic experiences of real-world object interactions. While speculative, we see parallels in the visuomotor patterns employed by prosthetic limb users (Cheng et al., 2022), including those performing Lavoie et al.'s (2018) identical real-world task (Hebert et al., 2019), where a lack of haptic feedback is compensated by a reliance on visual feedback, perhaps extending the time spent looking at the drop-off site.

In the context of digital interactions, this work adds to the smaller set of works that explore the dynamics of visuomotor behaviours mediated by digital (e.g. screen-based) interactions. Our eye-leading-cursor results are broadly observed in other forms of digital interactions including shape tracing (Deng et al., 2016), distractorless visual search (Bieg et al., 2010), object tracking (Danion & Flanagan, 2018), simple (Sailer et al., 2000) and sequential (Rand & Stelmach, 2010) reaches towards a screen, and dragging virtual objects on a touchscreen (Sims et al., 2011). In these activities, this pattern appears across both cursor and touchscreen interactions, though these types of interactions typically result in smaller eye-lead times (within ∼100 to 300 ms). Our longer eye-lead times, despite being more in line with real-world timings (e.g. Hayhoe, 2000; Land and Hayhoe, 2001; Land et al., 1999; Lavoie et al., 2018), likely reflect how eye-cursor coordination is shaped by the nature of a specific digital interaction. That is, there is other evidence of specific task features affecting visuomotor patterns (Liebling & Dumais, 2014; van Donkelaar & Staub, 2000), and our drag and drop task has greater demands to accurately select, drag and release the object within the narrow bounds of the pick-up and drop-off areas. Conversely, the smaller eye-lead times were measured in digital tasks that were primarily simple reach, point, or click responses (as opposed to clicking-dragging-releasing), where less anticipatory looking ahead may be required.

Arriving at these types of conclusions took a lot of work. Unlike the above cited works that used lab-grade eye-trackers, our choice to employ webcam eye-tracking (for various reasons already discussed) meant we had to contend with its limitations. Beyond the hurdles in gathering the data (as recounted in Chapter 4), deriving meaning from the raw gaze data output from webcam eye-tracking proved to be our most daunting task. Despite our intentions to reproduce the metrics derived from real-world lab-grade eye-tracking (Lavoie et al., 2018), we were confronted with the realities of lower quality raw gaze data that, on the surface, appeared to render this goal unattainable. Simply following the same steps used by Lavoie et al. to convert the raw gaze data into our target metrics would have yielded nothing but meaningless noise. That is, trying to extract only instances of the raw gaze coordinates falling within the coordinate bounds of the pick-up/drop-off locations, for example, would have resulted in very limited spatial overlap. Thus, data from entire trials, and even entire participants, might lack instances of alignment between the raw gaze data and any of the four pick-up/drop-off locations.

Instead, constrained by the data quality afforded by webcam eye-tracking, we had to develop a solution to our noisy data quality problem. Visual inspection of the gaze and cursor behaviour over time revealed within-subject spatial distortions in the gaze data, yet their movements seemed to align in direction and timing. Put another way, the gaze data seemed to follow the task in time but was distorted in space. Recognizing that there was potentially something to be salvaged, we developed a clustering approach for the skewed gaze data, using the cluster centroids as the center of the 4 AOIs (assigned based on their relative spatial orientation), and then binned the raw gaze data to its nearest AOI/centroid. This transformation addressed the skewed raw gaze data by rendering the absolute (and distorted) spatial location irrelevant. The cost, however, was a significantly diminished spatial resolution, with sensitivity reduced to only these 4 AOIs. Moreover, this strategy didn't resolve all of our data quality concerns; by the end, we had to exclude 22 out of the 51 participant datasets from analysis due to issues with eye data quality. This example of data processing is another testament to the tradeoffs we faced when adopting webcam eye-tracking; it allowed us to successfully track how the eyes led the cursor, but it hindered our exploration of certain real-world metrics from Lavoie et al. (2018). For example, without the spatial sensitivity to track when the gaze and cursor intersected, we were unable to see whether the change from a real-world hand to a mouse-cursor affected effector-gaze interactions (like in Lavoie et al., 2018). More specifically, the previous real-world results also reported the amount of time spent looking at the hand, an important indicator of task proficiency but something we couldn't even conceptually replicate with this data. Finally, the data processing decision we made to reduce the gaze data to 4 AOIs not only exposed the limits of our chosen eye-tracking method but also made it necessary to corroborate our empirical results with additional supportive "proofs". Being transparent, reviewers demanded that we show evidence beyond the main empirical findings that the data quality was sufficient to base our results on. As such, we included demonstrations that the expected gaze behaviour pattern throughout a trial followed functional predictions (i.e. greater fixation time on relevant AOIs compared to irrelevant AOIs, see Chapter 2, Figure 2.3) and we performed a spatial probability analysis to confirm the high likelihood of our clustering approach accurately assigning gaze data to the appropriate AOI.

5.1.2 Chapter 3 - Assessing webcam eye-tracking utility in binary choice decision-making

Chapter 3 was strategically designed to evaluate the utility of webcam eye-tracking within a decision-making context. Our primary objective was to replicate and extend the results of the first paper in this two-part series: a binary choice investigation conducted by Ouellette Zuk et al. (2023; included in the Appendix). The binary choice decisions spanned diverse decision domains and featured a range of stimuli, all presented in the same layout. With the participation of 100 remote, crowdsourced participants, three binary choice tasks - Sentence Verification (Dale & Duran, 2011), Numeric Size Congruity (Faulkenberry et al., 2016), and Photo Preference (Koop & Johnson, 2013) - were tested. Our approach to evaluating webcam eye-tracking's utility was to explore whether it could enhance our understanding of decision difficulty, a factor robustly illustrated in Ouellette Zuk et al.'s cursor movement measures. Rather than adopting a specific, hypothesis-driven approach as in Chapter 2, Chapter 3 was more exploratory. That is, in Chapter 2 we knew what effect we predicted and what measure we'd use to chart it, but in Chapter 3, while we knew the effect of interest (decision difficulty) its specific manifestation in gaze behaviour was an open question. By using webcam eye-tracking to capture gaze data, we progressed beyond cursor movement measures, uncovering decision difficulty effects earlier in time. Specifically, we unveiled a pattern of increased gazes at the choice options for hard trials than easier ones, an effect consistently observed across all three binary choice tasks.

Yet, webcam eye-tracking facilitated an even deeper exploration of nuanced decision-difficulty effects by revealing task-specific gaze patterns that mirrored the task-specific spatial distribution of decision-relevant information. Despite the uniform presentation format of the binary decision tasks, varying decision types necessitated distinct approaches to acquiring the decision-relevant information. For instance, choosing a preferred photo required scrutinizing both options, with minimal attention to the repetitive task instructions at the top of the screen. In contrast, to verify the truthfulness of a sentence, the sentence needed to be read, while peripheral vision may have been sufficient for indicating a decision via cursor placement due to the fixed 'true' and 'false' choice options. In the Numeric Size Congruity task, identifying the larger numeric (and not necessarily physical) number in two options is somewhere in the middle: akin to Photo Preference both choice options contain important information, but because of the stimuli set used, there were times that the correct answer could be deduced with only one option viewed. Webcam eye-tracking illuminated these variations in how information was spatially distributed, informing further task-specific analyses of the most prevalent gaze patterns. By dissecting these prevalent patterns, we uncovered novel difficulty effects in each task, tracing the decision-making process through the time spent looking at task-relevant information. In the end, webcam eye-tracking demonstrated its utility while simultaneously contributing to a deeper understanding of the decision-making process.

Positioning our findings within the broader literature context, we begin with the preceding (and accompanying; see Appendix A) paper that laid the foundation for this contribution (Ouellette Zuk et al., 2023). While this earlier study elegantly showcased the value of remote data collection for continuous cursor and touchscreen trajectories in decision-making, it grappled with a notable limitation: the inability to reveal pre-movement decision processes. This knowledge gap underscores the significance of Chapter 3, where webcam eye-tracking is employed to reveal information about the decision-making process that occurs prior to movement. Granted, we know that once the cursor starts moving, we can access insights about the decision making process, where decision difficulty is reflected in the duration and trajectory curvature of cursor movements (Faulkenberry et al., 2016; Freeman, 2018; Hehman et al., 2015; Koop & Johnson, 2013; Maldonado et al., 2019; Stillman et al., 2018, 2020) . However, we also know that the decision-making process is sensitive to decision difficulty *prior* to movement initiation, as evidenced by a robust history of response timing effects where harder decisions take longer than easier decisions (e.g. McCarthy and Donchin, 1981; Palmer et al., 2005; Rangel and Hare, 2010; Schouten and Bekker, 1967). Therefore, our Chapter 3 adds to a growing collection of works that have employed eye-tracking as a means of understanding the dynamics of decision-making (for reviews see Bhatnagar and Orquin, 2022; Orquin and Mueller Loose, 2013) - *how* decision-making unfolds across the entire decision, including the time before movement where eye movements are required to sample the relevant decision information.

Diving into these pre-movement mechanisms further, gaze patterns have been shown to both bias and reflect choices (Glaholt et al., 2009; Gold & Shadlen, 2007; Shimojo et al., 2003). Classic decision-making models and theories (e.g. evidence accumulation; Gold and Shadlen, 2007; Ratcliff and Rouder, 1998; P. L. Smith and Vickers, 1988) have incorporated the important role gaze plays in information sampling (e.g. the attentional drift diffusion model, aDDM; Krajbich et al., 2010), where the eyes act as an accumulator of information samples for different alternatives until enough information is accumulated to make a choice (Krajbich & Rangel, 2011; Krajbich et al., 2010). The aDDM has primarily been modeled after decisions like our Photo Preference task, where gaze patterns during preferential binary choices are predicted to begin with the gaze on the left option and conclude with the gaze on the chosen option, with harder choices requiring more and longer choice option fixations (Krajbich et al., 2010). Remarkably, despite the various constraints imposed by webcam eye-tracking (discussed later), our analysis of the most common gaze patterns during Photo Preference aligns with the aDDM, and to the best of our knowledge, is the first time remotely-captured webcam eye-tracking has revealed these effects.

Despite the aDDM being extended and tested with different stimuli types (e.g. numeric vs pictorials; Krajbich et al., 2012) and choice domains (e.g. risk and social choices; S. M. Smith and Krajbich, 2018), it wasn't clear how gaze patterns would express decision difficulty for our Sentence Verification and Numeric Size Congruity tasks. However, our results offer interesting complementary evidence for the role gaze plays in information gathering. Previous work has shown that, during three-alternative preferential choices, only after all three choices were viewed did the underlying preferential value of a choice affect the fixation process (Krajbich & Rangel, 2011). While its carryover into non-preference choices is speculative, our work supports the notion that early fixations reflect only information gathering (and not choice discrimination), with decision difficulty effects emerging only later in time (Krajbich & Rangel, 2011). However, perhaps the greatest and most novel offering of Chapter 3 comes from the powerful message it can provide when considered alongside its counterpart (Ouellette Zuk et al., 2023): the combination of cursor-tracking *and* webcam eye-tracking produces an immense opportunity to understand the entire timeline of a decision - from stimulus onset to the end of action enactment - all while captured entirely remotely. Together, these works demonstrate to decision-making researchers that it's possible to measure a complicated human process beyond the confines of the laboratory, enabling the capture of more authentic and realistic decision-making.

The ability to remotely explore the complexities of the decision making process across time is exciting, but much like Chapter 2, it was a challenge to arrive at meaningful measures that could offer these insights. Chapter 3 presented a novel challenge. Unlike Chapter 2's focus on finding ways to apply established and specific gaze-cursor coordination metrics to noisy webcam eye-tracking data, Chapter 3 did not have nearly the same specificity of target metrics. Instead, our process to generate metrics was different - we had some idea from gaze-aware decision-models that gaze pattern features like where, how often, and for how long you looked would be important (Krajbich & Rangel, 2011; Krajbich et al., 2010), but we also knew that webcam eye-tracking imposed spatial and temporal limitations that wouldn't give us the power to generate the same kind of highly-precise saccade-based gaze metrics that were foundational to models like the aDDM (Krajbich et al., 2010). Thus, akin to Chapter 2, our first step towards meaningful measures was a spatial dimensionality reduction to the raw gaze data, assigning the gaze data into one of the two choice option AOIs if its coordinates fell within the moderately-expanded AOI boundaries.

Likely due to improvements in the quality of our chosen webcam eye-tracking program (which was under continuous development throughout both studies), our initial visual inspections and coarse metrics (e.g. looking at the number of trials where the raw gaze data never fell within either choice option) indicated that more advanced data treatments (like the clustering approach used to resolve skewed gaze data in Chapter 2) were unnecessary. However, during these early data explorations, we quickly came to realize that despite the task's outward appearance of being highly constrained (i.e. only two choice options in a simple decision paradigm involving easy and hard trials), we needed to contend with the unconstrained nature of exploratory eye movements. Unlike a cursor, which must land on a specific choice option for the task to proceed, eyes sampling the choice options possess the freedom to move without any task-driven constraints. By that I mean that despite the aDDM, for example, suggesting that the first gaze during decision-making falls on the left choice option, in reality, not all participants looked at the left option first in every trial for each task. To address this, our approach involved surveying features of the gaze data as a means of building a picture of the most commonly-employed gaze strategies. This is how we arrived at our within-task proportion of trials measures, where the use of a counting measure meant we could test the full range of features (e.g. like 1 to 4 dwells) without putting ourselves at risk of a failed statistical test due to empty cells (i.e./e.g. where most participants never had a trial with 4 dwells). This proportion analysis approach uncovered the substantial differences in gaze patterns between tasks, and informed the further lines of analyses that explored the dynamics of the most common gaze patterns. For clarity, the unconstrained nature of eye data is a problem in all eye-tracking contexts, but introduces unique challenges with the dimensionality reduction demanded by webcam eye-tracking data.

5.1.3 Chapter 4 - A practical guide to webcam eye-tracking

Our fourth chapter was a wholly methodological contribution, offering a wealth of practical insights and empirical context for conducting webcam eye-tracking experiments. Within Chapter 4, we deeply explored the various considerations we encountered and tested during Chapters 2 and 3, presenting a comprehensive guide on how to maximize the potential of webcam eye-tracking. By employing real, remotely-captured experimental data, we untangled the method's limitations and anticipated benefits, illuminating the challenges associated with participation (including participation requirements and setting system thresholds), the expected quality of gaze data (both temporally and spatially), and the prospective costs of running webcam eye-tracking studies. Our fourth chapter goes beyond a technical validation of the method, and instead addresses a wide range of practical considerations in order to minimize the method's barrier to entry for aspiring webcam eye-tracking researchers.

The primary aim of Chapter 4 was to share our knowledge with prescriptive information, thereby enhancing the utility of webcam eye-tracking as a research method. To the best of my knowledge, no other publication offers this kind of comprehensive guidance for navigating the intricacies of webcam eye-tracking experimentation. Previously, this kind of information was fragmented, scattered across the discussion or limitations sections within the limited collection of webcam eye-tracking papers. Now, we've addressed this gap, offering an organized collection of practical information in a guiding resource, in an effort to encourage the greater adoption of webcam eye-tracking, and ultimately enhance its utility. This paper addresses this gap and ideally will streamline the experiences of future webcam eye-tracking researchers, fostering greater efficiency and satisfaction in their endeavors.

5.2 What is utility anyways?

Throughout this thesis, I've extensively explored and discussed the utility of webcam eye tracking. When I've talked about its utility, I've described it in terms of its usefulness – if it adequately serves as a tool to measure certain gaze behaviours or if it empowers us to make specific assertions. This was how utility was discussed in Chapters 2 and 3, but in the subtext of these studies, and when more deeply interrogated in Chapter 4, it becomes clear that a method's utility is defined by much broader terms.

In practice, evaluating the utility of any tool, including webcam eye-tracking, should not happen in a vacuum. Rather, its utility must be contextualized within a broader framework that encompasses practical aspects like efficiency, user-friendliness, and costs. In fact, it was this more broad and holistic form of utility that I unknowingly assessed when I faced the challenge of continuing my PhD research during the pandemic-induced shift to remote work. In early April 2020, as in-person research was stalled, I scoured the internet for remote-friendly methods to proceed with my eye-tracking research. Encountering options like Webgazer (an open-source offering from Papoutsaki et al., 2016), which I discussed in the Introduction (Chapter 1), was underwhelming as I came to realize that its temporal and spatial accuracy was not sufficient to let me ask the kinds of research questions I was interested in answering. Importantly, this utility was a reflection of my needs as much as it was a judgment of the tool - many others have found Webgazer sufficient for their purposes (e.g. Semmelmann and Weigelt, 2018; Slim and Hartsuiker, 2022; Vos et al., 2022; Yang and Krajbich, 2021), speaking to how much of the utility is user-dependent.

For example, my evaluation of utility necessarily went beyond a data quality assessment; I also considered whether I could effectively implement unfamiliar JavaScript tools in an online experiment. For some with web development experience, immense utility might be derived from a free open-source webcam eye-tracker like Webgazer (Papoutsaki et al., 2016), providing the flexibility to optimize and modify the eye-tracking algorithm (e.g. Yang and Krajbich, 2021). However, for me and many researchers wanting to conduct online studies, practical utility is limited to user-friendly tools that don't require degrees in computer science. Thus, my preference shifted to Labvanced (Finger et al., 2017), an alternative offering that employed advanced deep neural network algorithms and included an entire online experiment-building ecosystem. Actual utility in this case was achieved both because the data quality was sufficient and because of the sufficiently intuitive programming interface. Nevertheless, as detailed throughout this thesis and highlighted in Chapter 4, this alternative still presented its own challenges, necessitating considerations of its constraints and data processing requirements.

Like my assessment of Webgazer's utility, others may evaluate my chosen webcam eye-tracking approach based on their ability to navigate its constraints. For some, like a resource-strained business, the limitations of the current methods may make it incredibly difficult to derive utility from webcam eye-tracking. In a business context, without an eye-tracking expert on staff it would only be a matter of days before an unaffordable return on investment of webcam eye-tracking was realized. However, in my case, prior experience handling noisy human timeseries data (i.e. EEG, motion capture, and in-lab eye-tracking data) eased the data processing aspects, and the academic context allowed for dedicated time to overcome these hurdles (so much so that it became a central theme of this very thesis!). Therefore, the user's expertise and abilities, especially as they relate to handling methodological challenges or constraints, are also important factors when considering the utility of a method.

Ultimately, webcam eye-tracking's utility is in the eye, and expertise, of the experimenter (or other interested beholders). Utility is contingent on what questions need to be answered or what outcomes need to be delivered by the tool *as well as* the user's flexibility and willingness to operate or work through the method's imposed constraints. In turn, in addition to judging what use a given tool like webcam eye-tracking can provide, utility is also dictated by the expertise of the user - to wit almost everyone can pick up a hockey stick, but almost no one can play hockey like Connor McDavid. In the end, we need to take a comprehensive view of utility, including for webcam eye-tracking and recognize that this judgment resonates with a major theme of this thesis - there are tradeoffs to be considered both with respect to the demands we place on our tools and on ourselves.

5.3 The future of webcam eye-tracking

At the very start of this thesis, I painted a picture of the future, where robot pups and eye-tracking were a part of enriching our everyday lives. That picture was a rosy one, and for the most part, this thesis has continued to paint a rosy image of webcam eye-tracking. However, it's easy (for me, at least) to get swept up in all the excitement. I'm often reminded of the more realistic challenges for webcam eye-tracking's future when I'm asked by family or friends about this thesis. The typical response to a thesis about webcam eye-tracking is usually something like "Oh! Wow... like spying on people through their little computer screen cameras?!". Of course, I provide some important clarification, but it's difficult to fully relieve their hesitations. I don't blame anyone for having these concerns - we've witnessed massive violations of user privacy from once-trusted technology companies (e.g. Facebook and Cambridge Analytica; Isaak and Hanna, 2018), and efforts to protect consumers with privacy regulations consistently fail to keep pace with emerging technologies (Hacker et al., 2023). In fact, a recent global survey by the International Association of Privacy Professionals (IAPP) found that 68% of the 4750 consumers surveyed were either somewhat or very concerned about their privacy online (*IAPP Privacy and Consumer Trust Report*, 2023).

Privacy concerns are further heightened when sensitive biometric data is at stake. In May 2023, the American Federal Trade Commission (FTC) issued a policy statement to address the high risk of misuse of biometric data technologies like fingerprint, iris, and facial recognition (*Policy Statement of the FTC on Biometric Information and Section 5 of the FTC Act*, 2023). The use of camera-captured biometric data for facial recognition has received particular scrutiny, especially for its pervasive use in law enforcement (e.g. half of the American adult population are in a law enforcement facial recognition network; Georgetown Law Center on Privacy & Technology, Garvie et al., 2016). This has raised concerns among public privacy advocates, particularly alarmed by the potential for harm when facial recognition technology, with known racial biases (e.g. significantly more false positive errors for West and East African and East Asian faces than for Eastern European faces; Duewer, 2022), is deployed in real-time with body-worn cameras (Ringrose, 2019). Given this current technology and privacy ecosystem, it begs the question how we can imagine anything besides a bleak future for a tool like webcam eye-tracking. This next section discusses the pessimist's view of webcam eye-tracking's future, but also the optimist's perspective - the one I hope you, the reader, will share with me.

5.3.1 The pessimist's perspective

While the academic sphere benefits from established ethical practices and guidelines, concerns arise when envisioning the application of webcam eye tracking beyond academia. Mishandled implementation of this technology in non-academic contexts could pose significant challenges for society at large. In particular, webcam eye-tracking's integration with industrial purposes would demand meticulous consideration, transparent explanations, and consumer awareness. This concern emerges in a broader technological landscape where transparency around technology is increasingly diminished, with artificial intelligence applications that make use of large language models (e.g. ChatGPT) being an obvious example of this today. This lack of transparency is coupled with the rising apprehension surrounding the rapidly growing Internet of Things and technology's omnipresence in our lives (Isaak & Hanna, 2018). Therefore, it comes as no surprise that the use of webcam technology is uncomfortable for some. In fact, looking around a coffee shop or a university library, it's incredibly common to see laptops with makeshift covers on the device-integrated webcams. Amidst this landscape, worries about user privacy and security intensifies, including fears of unauthorized camera access. Moreover, the use of computing devices has only become more ubiquitous, interwoven with our daily activities and environments, ensuring their constant proximity to us. This setting harbors both opportunities, as I will explore from an optimistic perspective, and challenges, including the potential misuse of this technology by ill-intentioned actors.

If we think back to the eye-tracking future envisioned in the Introduction (Chapter 1), it's not hard to fathom more dangerous uses of eye-tracking data that aren't as innocent as the suggestion of a smoothie based on your fridge perusing. Chapter 3 demonstrated the (exciting) utility of webcam eye-tracking for understanding decision-making processes, but there's plenty of cases where you might not want third parties privy to the ease or difficulty you experienced making a decision. Using a sobering yet plausible example, imagine your gaze data being deceptively captured while browsing on your own laptop, contemplating your options pertaining to a highly-politicized (and in some places illegal) personal health choice. The notion that your extended and repeated fixations on the “Book Your Appointment Now” button could be exploited to discern your struggle in arriving at this private choice is unsettling. This becomes particularly alarming if such information were employed to detect your susceptibility or receptiveness to the persuasions of organizations with histories of dangerous, unethical, and exploitative initiatives. Even if webcam eye-tracking data is technically collected with user consent, the current landscape is unfortunately friendly to this kind of future. This is particularly concerning for vulnerable populations, who may fall victim to opaque attempts to satisfy user consent regulations with ‘accept all cookies’ buttons and impenetrable terms and conditions, unknowingly sharing their gaze data with ill-intentioned parties.

5.3.2 The optimist’s perspective

All technology is susceptible to bad actors and manipulative uses, but I believe the glass is at least half full when it comes to webcam eye-tracking. In that future, webcam eye-tracking as a technology will see exciting advancements, evolving to include all forms of consumer cameras, as has been seen with recent advances in smartphone eye-tracking (e.g. Khamis et al., 2018; Le et al., 2022; Namnakani et al., 2023; Valliappan et al., 2020). With the promise of the Apple Vision Pro, and with current virtual reality headsets, it’s clear that mobile, lightweight, and consumer-friendly eye-tracking technology is here - and only in its infancy. This is exciting for many reasons - it makes the technology accessible, and encourages its portability to domains previously deemed unfeasible for applications of traditional eye-tracking tools. For example, recent feasibility studies have explored the use of webcam eye-tracking with populations like older adults with Alzheimer’s Disease (Greenaway et al., 2021), infants (Bánki et al., 2022), and students in a neurodiverse classroom

(Wong et al., 2023). It's exciting to imagine the various instances where webcam eye-tracking could be (or continue to be) used for good: as a powerful assistive technology tool that can offer alternative, more accessible options for computer interaction (e.g. eye-gaze or gaze-assistive technology; Donegan et al., 2009; Lariviere, 2015), as a portable and practical diagnostic tool that could be used, for example, in a sport context to perform real-time concussion assessments, or as a telehealth measurement tool to remotely quantify and monitor rehabilitation efficacy after a traumatic brain injury or stroke.

A future with webcam eye-tracking offers many opportunities to benefit the common good. While it's perhaps a more nuanced example, our demonstration of the utility of webcam eye-tracking in understanding digital object interactions illuminates opportunities for better, and potentially safer, interaction design. Imagine a scenario where efficient software use was a matter of life and death, like in disaster or emergency response settings. Testing this software on the jobsite with webcam eye-tracking for the presence of the $\sim$500 ms look-ahead gaze pattern might reveal opportunities for critically improving the interaction design (e.g. adjustments to object sizing or positioning, or incorporation of haptic feedback). As another example, the presence of this pattern as an index of good interaction design might also be meaningful for designing adaptive and accessible user experiences. Specifically, with webcam eye-tracking, one could imagine a future where a user interface would dynamically adapt its features in an optimization effort to achieve a near-500 ms look-ahead gaze pattern. This would be particularly meaningful for software that serves a critical purpose and diverse populations (e.g. tax preparation software), or for software with a user base that relies extensively on it for their everyday job duties (e.g. specialized B2B software). Beyond these applications, all three manuscripts, and especially Chapter 4, serve to encourage and support further webcam eye-tracking research, contributing to a foundation and ecosystem that encourages the application of webcam eye-tracking technology to provide innovative solutions that can help address important issues like the accessibility of technology, workplace safety, and health diagnostics.

5.3.3 The realist's perspective

In their review entitled "What Does Your Gaze Reveal About You? On the Privacy Implications of Eye Tracking", Kröger et al. (2020) write:

"According to the reviewed literature, eye tracking data may reveal information about a user's biometric identity, mental activities, personality traits, ethnic background, skills and abilities, age and gender, personal preferences, emotional state, degree of sleepiness and intoxication, and physical and mental health condition."

After discussing webcam eye-tracking with both pessimism and optimism, if it weren't for the title, this excerpt from Kroger et al.'s (2020) review could easily resonate with either perspective. In fact, it helped me clarify my own more optimistic (and perhaps naive) perspective as, despite reading it in-text, surrounded by the authors' alarm of eye-tracking's "serious privacy concerns", it only made me excited to see the breadth of information eye-tracking affords us. Even if you, the reader, don't share in my (naive) optimism, it's not hard to envision good and ethical use cases for most of these qualities, with the potential of great benefits for the end user. At the same time, the use of this data without the end user's consent, or for commercial purposes that only serve corporate shareholder objectives, is admittedly terrifying.

In reality, both sides of this coin exist, and this thesis doesn't aim to resolve this tension. Instead, its purpose is to illuminate both the possibilities and the inevitabilities as we navigate a future of webcam eye-tracking technology. As transformative applications arise, so will instances of misuse, and sometimes, it will be difficult to give absolute or stable assignments of "good" or "bad" to the various uses of webcam eye-tracking technology. During my industry internship, I saw first hand just how close we are to the broad use of webcam eye-tracking for commercial applications. In fact, a commercial tool rooted in Chapter 3's cursor-tracking predecessor (Ouellette

Zuk et al., 2023) already exists today, using decision difficulty indices from cursor trajectories for pre-interview job candidate screening, influencing hiring practices. These sorts of tools clearly have life-changing implications, but in reality, their creators are likely unaware or ignorant to the full scope of their tool's consequences. In some ways, my thesis has contributed to this problem, unintentionally unaware of all the possible ramifications tied to webcam eye-tracking's development. Nevertheless, this thesis can also enlighten - it demonstrates the availability and utility of this technology, and now, offers a realistic perspective on the good and the bad that come with webcam eye-tracking.

5.4 My final thoughts: a reflection

I am very excited about the power of webcam eye-tracking, and its potential for learning more about human behaviour and contributing to a more accessible and efficient future. The challenges that I had to resolve while working my way through Chapter 2 and 3, and even prior, as I transitioned my research from the lab to the internet, seemed nearly impossible to overcome at the time. With little availability of supportive resources, it took about one year to end up with my first complete webcam eye-tracking dataset. Now, in retrospect, some of the challenges are no longer relevant, totally dissolved by technological improvements or system advancements. Taking a more big-picture view of it all, it's likely that a lot of the limitations of webcam eye-tracking that I worked through are going to lose their relevance given the accelerated pace technology development moves at. This is exciting and daunting all at the same time, as it only means new challenges will present themselves. Yet, however the method evolves, its utility will only continue to grow, and for that I remain excited and feel grateful to have made a contribution in service of that mission.

In the end, I hope these works have convinced you of the power of webcam eye-tracking. In object interactions, decision-making, or other domains, webcam eye-tracking offers a lot, and with each methodological contribution, like Chapter 4, the return on investment grows for researchers. It has been rewarding (albeit frustrating at times) to dive deeply into a new method. There was no guidebook or manual to this when I first embarked on this journey. Instead, my interest in exploring the limits of webcam eye-tracking resulted in becoming an 'advanced user' of our chosen webcam eye-tracking tool (Labvanced), where my ongoing contact with the system's creators evolved into a rewarding year-long internship experience with their company. Being an 'advanced user' might be called 'cutting edge', like my supervisor describes it, but understanding and working with this new method of eye-tracking was much less glamorous that what I imagined the cutting edge landscape to be. Instead, it involved working in the murky, under-defined frays, trying to make use of a tool that was still itself being built. Its novelty, and perhaps also its confinement to the online space made this experience particularly isolating, unlike my past experiences of navigating a tool like EEG with my lab mates, other supportive EEG labs on campus, online wikis and forums, or the manufacturer's scientific consultants. But, despite this, it's exactly this novelty and online accessibility that has me so excited for the future of this method. Moving beyond the laboratory will be a defining feature of this modern human research era, and finding ways to do this with the most effective tools is crucial. There is still a lot more to learn, but with eye-trackers (i.e. cameras) built into most computers and smartphones, we're well on our way to more deeply understanding everyday human behaviour.

5.5 References

Bánki, A., de Eccher, M., Falschlehner, L., Hoehl, S., & Markova, G. (2022). Comparing online webcam- and laboratory-based eye-tracking for the assessment of infants' audio-visual synchrony perception. *Frontiers in Psychology*, *12*, 733933.

Bhatnagar, R., & Orquin, J. L. (2022). A meta-analysis on the effect of visual attention on choice. [Publisher: US: American Psychological Association]. *Journal of Experimental Psychology: General*, *151*(10), 2265.

Bieg, H.-J., Chuang, L. L., Fleming, R. W., Reiterer, H., & Bülthoff, H. H. (2010). Eye and pointer coordination in search and selection tasks. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10*, 89.

Butsch, R. L. C. (1932). Eye movements and the eye-hand span in typewriting [Place: US Publisher: Warwick & York]. *Journal of Educational Psychology*, *23*(2), 104–121.

Cheng, K. Y., Chapman, C. S., & Hebert, J. S. (2022). Spatiotemporal coupling of hand and eye movements when using a myoelectric prosthetic hand [ISSN: 1945-7901]. *2022 International Conference on Rehabilitation Robotics (ICORR)*, 1–6.

Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science*, *35*(5), 983–996.

Danion, F. R., & Flanagan, J. R. (2018). Different gaze strategies during eye versus hand tracking of a moving target. *Scientific Reports*, *8*(1), 10059.

Deconinck, F. J. A., van Polanen, V., Savelsbergh, G. J. P., & Bennett, S. J. (2011). The relative timing between eye and hand in rapid sequential pointing is affected by time pressure, but not by advance knowledge. *Experimental Brain Research*, *213*(1), 99–109.

Deng, S., Chang, J., Kirkby, J. A., & Zhang, J. J. (2016). Gaze–mouse coordinated movements and dependency with coordination demands in tracing. *Behaviour & Information Technology*, *35*(8), 665–679.

Donegan, M., Morris, J. D., Corno, F., Signorile, I., Chió, A., Pasian, V., Vignola, A., Buchholz, M., & Holmqvist, E. (2009). Understanding users and their needs. *Universal Access in the Information Society*, *8*(4), 259–275.

Duewer, D. L. (2022). *Face recognition vendor test (FRVT) part 8:: Summarizing demographic differentials* (NIST IR 8429). National Institute of Standards and Technology. Gaithersburg, MD.

Faulkenberry, T. J., Cruise, A., Lavro, D., & Shaki, S. (2016). Response trajectories capture the continuous dynamics of the size congruity effect. *Acta Psychologica*, *163*, 114–123.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: A unified JavaScript framework for online studies. *International Conference on Computational Social Science*, *1*(1), 1–3.

Freeman, J. B. (2018). Doing psychological science by hand [Publisher: SAGE Publications Inc]. *Current Directions in Psychological Science*, *27*(5), 315–323.

Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1436), 2435–2440. Retrieved September 13, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1690464/

Garvie, C., Bedoya, A. M., Frankle, J., Daugherty, M., Evans, K., George, E. J., McCubbin, S., Rudolph, H., Ullman, I., Ainsworth, S., Houck, D., Iorio, M., Kahn, M., Olson, E., Petenko, J., & Singleton, K. (2016). UNREGULATED POLICE FACE RECOGNITION IN AMERICA.

Glaholt, M. G., Wu, M.-C., & Reingold, E. M. (2009). Predicting preference from fixations [Place: Italy Publisher: PsychNology Journal]. *PsychNology Journal*, *7*, 141–158.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making [_eprint: https://doi.org/10.1146/annurev.neuro.29.051605.113038]. *Annual Review of Neuroscience*, *30*(1), 535–574.

Greenaway, A.-M., Nasuto, S., Ho, A., & Hwang, F. (2021). Is home-based webcam eye-tracking with older adults living with and without alzheimer's disease feasible? *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, 1–3.

Hacker, P., Engel, A., & Mauer, M. (2023). Regulating ChatGPT and other large generative AI models. *2023 ACM Conference on Fairness, Accountability, and Transparency*, 1112–1123.

Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, *7*(1), 43–64.

Hebert, J. S., Boser, Q. A., Valevicius, A. M., Tanikawa, H., Lavoie, E. B., Vette, A. H., Pilarski, P. M., & Chapman, C. S. (2019). Quantitative eye gaze and movement differences in visuomotor adaptations to varying task demands among upper-extremity prosthesis users. *JAMA Network Open*, *2*(9), e1911197.

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, *18*(3), 384–401.

IAPP privacy and consumer trust report. (2023, March). International Association of Privacy Professionals. Retrieved August 17, 2023, from https://iapp.org/media/pdf/resource_center/privacy_and_consumer_trust_report_summary.pdf

Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, cambridge analytica, and privacy protection [Conference Name: Computer]. *Computer*, *51*(8), 56–59.

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye–hand coordination in object manipulation. *The Journal of Neuroscience*, *21*(17), 6917–6932.

Khamis, M., Alt, F., & Bulling, A. (2018). The past, present, and future of gaze-enabled handheld mobile devices: Survey and lessons learned. *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–17.

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, *67*(4), 151–185.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology*, *3*. Retrieved May 25, 2023, from https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00193

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13852–13857.

Kröger, J. L., Lutz, O. H.-M., & Müller, F. (2020). What does your gaze reveal about you? on the privacy implications of eye tracking. In M. Friedewald, M. Önen, E. Lievens, S. Krenn, & S. Fricker (Eds.), *Privacy and identity management. data for better living: AI and privacy: 14th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2.2 international summer school, windisch, switzerland, august 19–23, 2019, revised selected papers* (pp. 226–241). Springer International Publishing.

Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, *25*(3), 296–324.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities. *Vision Research*, *41*(25), 3559–3565.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328.

Lariviere, J. A. (2015). Eye tracking: Eye-gaze technology. In I. Söderback (Ed.), *International handbook of occupational therapy interventions* (pp. 339–362). Springer International Publishing.

Lavoie, E. B., Valevicius, A. M., Boser, Q. A., Kovic, O., Vette, A. H., Pilarski, P. M., Hebert, J. S., & Chapman, C. S. (2018). Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. *Journal of Vision*, *18*(6), 1–20.

Le, T., Dietz, F., Pfeuffer, K., & Alt, F. (2022). A practical method to eye-tracking on the phone: Toolkit, accuracy and precision. *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia*, 182–188.

Liebling, D., & Dumais, S. (2014). Gaze and mouse coordination in everyday work. *UBICOMP ADJUNCT '14*, 10.

Maldonado, M., Dunbar, E., & Chemla, E. (2019). Mouse tracking as a window into decision making. *Behavior Research Methods*, *51*(3), 1085–1101.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of p300 latency and reaction time [Publisher: American Association for the Advancement of Science]. *Science*, *211*(4477), 77–80.

Namnakani, O., Abdrabou, Y., Grizou, J., Esteves, A., & Khamis, M. (2023). Comparing dwell time, pursuits and gaze gestures for gaze interaction on handheld mobile devices. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–17.

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206.

Ouellette Zuk, A. A., Bertrand, J. K., & Chapman, C. S. (2023, June 7). Continuous measures of decision-difficulty captured remotely: I. mouse-tracking sensitivity extends to tablets and smartphones [Pages: 2023.06.06.543796 Section: New Results].

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*(5), 1.

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions.

Patla, A., & Vickers, J. (2003). How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental Brain Research*, *148*(1), 133–138.

Pelz, J., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, *139*(3), 266–277.

Policy statement of the FTC on biometric information and section 5 of the FTC act (Policy Statement). (2023). Federal Trade Commission. Retrieved August 25, 2023, from https://www.ftc.gov/system/files/ftc_gov/pdf/p225402biometricpolicystatement.pdf

Rand, M. K., & Stelmach, G. E. (2010). Effects of hand termination and accuracy constraint on eye–hand coordination during sequential two-segment movements. *Experimental Brain Research*, *207*(3), 197–211.

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology*, *20*(2), 262–270.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions [Place: United Kingdom Publisher: Blackwell Publishing]. *Psychological Science*, *9*, 347–356.

Ringrose, K. (2019). Law enforcement's pairing of facial recognition technology with body-worn cameras escalates privacy concerns. *Virginia Law Review Online*, *105*, 57. https://heinonline.org/HOL/Page?handle=hein.journals/inbrf105&id=57&div=&collection=

Sailer, U., Eggert, T., Ditterich, J., & Straube, A. (2000). Spatial and temporal aspects of eye-hand coordination across different tasks. *Experimental Brain Research*, *134*, 163–173.

Schouten, J. F., & Bekker, J. A. M. (1967). Reaction time and accuracy. *Acta Psychologica*, *27*, 143–153.

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451–465.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317–1322.

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2011). Adaptive allocation of vision under competing task demands. *Journal of Neuroscience*, *31*(3), 928–943.

Slim, M. S., & Hartsuiker, R. J. (2022). Moving visual world experiments online? a web-based replication of dijkgraaf, hartsuiker, and duyck (2017) using PCIbex and WebGazer.js. *Behavior Research Methods*.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*(2), 135–168.

Smith, S. M., & Krajbich, I. (2018). Attention and choice across domains. *Journal of Experimental Psychology: General*, *147*(12), 1810–1826.

Stillman, P. E., Krajbich, I., & Ferguson, M. J. (2020). Using dynamic monitoring of choices to predict and understand risk preferences [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *117*(50), 31738–31747.

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences*, *22*(6), 531–543.

Valliappan, N., Dai, N., Steinberg, E., He, J., Rogers, K., Ramachandran, V., Xu, P., Shojaeizadeh, M., Guo, L., Kohlhoff, K., & Navalpakkam, V. (2020). Accelerating eye movement research via accurate and affordable smartphone eye tracking. *Nature Communications*, *11*(1), 4553.

van Donkelaar, P., & Staub, J. (2000). Eye-hand coordination to visual versus remembered targets. *Experimental Brain Research*, *133*(3), 414–418.

Vos, M., Minor, S., & Ramchand, G. C. (2022). Comparing infrared and webcam eye tracking in the visual world paradigm [Accepted: 2023-01-03T13:11:04Z Publisher: University of California Press].

Wong, A. Y., Bryck, R. L., Baker, R. S., Hutt, S., & Mills, C. (2023). Using a webcam based eye-tracker to understand students' thought patterns and reading behaviors in neurodivergent classrooms. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 453–463.

Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, *16*(6), 1485–1505. Retrieved August 24, 2022, from https://journal.sjdm.org/21/210525/jdm210525.html

Bibliography

Abbott, W. W., & Faisal, A. A. (2012). Ultra-low-cost 3d gaze estimation: An intuitive high information throughput compliment to direct brain–machine interfaces [Publisher: IOP Publishing]. *Journal of Neural Engineering*, *9*(4), 046016.

Abdrabou, Y., Khamis, M., Eisa, R. M., Ismail, S., & Elmougy, A. (2019). Just gaze and wave: Exploring the use of gaze and gestures for shoulder-surfing resilient authentication. *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 1–10.

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations [Publisher: SAGE Publications Inc]. *Journal of Management*, *47*(4), 823–837.

Almaatouq, A., Becker, J., Houghton, J. P., Paton, N., Watts, D. J., & Whiting, M. E. (2021). Empirica: A virtual lab for high-throughput macro-level experiments. *Behavior Research Methods*, *53*(5), 2158–2171.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407.

Arora, K. (2019). Eye-head-hand coordination during visually guided reaches in head-unrestrained macaques. *Journal of Neurophysiology*, *122*, 1946–1961.

Bäck, D. (2006). *Neural network gaze tracking using web camera*. Institutionen för medicinsk teknik. Retrieved June 6, 2023, from https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-5579

Bader, F., Baumeister, B., Berger, R., & Keuschnigg, M. (2021). On the transportability of laboratory results [Publisher: SAGE Publications Inc]. *Sociological Methods & Research*, *50*(3), 1452–1481.

Balietti, S. (2017). nodeGame: Real-time, synchronous, online experiments in the browser. *Behavior Research Methods*, *49*(5), 1696–1715.

Bánki, A., de Eccher, M., Falschlehner, L., Hoehl, S., & Markova, G. (2022). Comparing online webcam- and laboratory-based eye-tracking for the assessment of infants' audio-visual synchrony perception. *Frontiers in Psychology*, *12*, 733933.

Bell, C. (1823). XV. on the motions of the eye, in illustration of the uses of the muscles and nerves of the orbit [Publisher: Royal Society]. *Philosophical Transactions of the Royal Society of London*, *113*, 166–186.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's mechanical turk [Publisher: Cambridge University Press]. *Political Analysis*, *20*(3), 351–368.

Bertrand, J. K., & Chapman, C. S. (2023). Dynamics of eye-hand coordination are flexibly preserved in eye-cursor coordination during an online, digital, object interaction task. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–13.

Bertrand, J. K., Zuk, A. A. O., & Chapman, C. S. (2023, June 7). Continuous measures of decision-difficulty captured remotely: II. webcam eye-tracking reveals early decision processing [Pages: 2023.06.06.543799 Section: New Results].

Bhatnagar, R., & Orquin, J. L. (2022). A meta-analysis on the effect of visual attention on choice. [Publisher: US: American Psychological Association]. *Journal of Experimental Psychology: General*, *151*(10), 2265.

Bieg, H.-J., Chuang, L. L., Fleming, R. W., Reiterer, H., & Bülthoff, H. H. (2010). Eye and pointer coordination in search and selection tasks. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10*, 89.

Bogdan, P. C., Dolcos, S., Buetti, S., Lleras, A., & Dolcos, F. (2023). Investigating the suitability of online eye tracking for psychological research: Evidence from comparisons with in-person data using emotion–attention interaction tasks. *Behavior Research Methods*.

Bowman, M. C., Johannson, R. S., & Flanagan, J. R. (2009). Eye–hand coordination in a sequential target contact task. *Experimental Brain Research*, *195*(2), 273–283.

Brand, A., & Bradley, M. T. (2012). Assessing the effects of technical variance on the statistical outcomes of web experiments measuring response times. *Social Science Computer Review*, *30*(3), 350–357.

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing megastudy: Comparing a range of experiment generators, both lab-based and online [Publisher: PeerJ Inc.]. *PeerJ*, *8*, e9414.

Bruno, A., Tliba, M., Kerkouri, M. A., Chetouani, A., Giunta, C. C., & Çöltekin, A. (2023). Detecting colour vision deficiencies via webcam-based eye-tracking: A case study. *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, 1–2.

Brunyé, T. T., Drew, T., Weaver, D. L., & Elmore, J. G. (2019). A review of eye tracking for understanding and improving diagnostic interpretation. *Cognitive Research: Principles and Implications*, *4*(1), 7.

Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods*, *50*(6), 2586–2596.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? [Publisher: SAGE Publications Inc]. *Perspectives on Psychological Science*, *6*(1), 3–5.

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of amazon's mechanical turk, its rapid rise, and its effective use [Publisher: SAGE Publications Inc]. *Perspectives on Psychological Science*, *13*(2), 149–154.

Burton, L., Albert, W., & Flynn, M. (2014). A comparison of the performance of webcam vs. infrared eye tracking technology. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1437–1441.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment [Place: US Publisher: American Psychological Association]. *Psychological Review*, *100*, 432–459.

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology and perception in art* [Pages: 198]. Univ. Chicago Press.

Butsch, R. L. C. (1932). Eye movements and the eye-hand span in typewriting [Place: US Publisher: Warwick & York]. *Journal of Educational Psychology*, *23*(2), 104–121.

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49–62.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? a comparison of participants and data gathered via amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156–2160.

Cassey, T. C., Evens, D. R., Bogacz, R., Marshall, J. A. R., & Ludwig, C. J. H. (2013). Adaptive sampling of information in perceptual decision-making [Publisher: Public Library of Science]. *PLOS ONE*, *8*(11), e78993.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among amazon mechanical turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112–130.

Chapman, C. S., Gallivan, J. P., Wood, D. K., Milne, J. L., Culham, J. C., & Goodale, M. A. (2010a). Reaching for the unknown: Multiple target encoding and real-time decision-making in a rapid reach task. *Cognition*, *116*(2), 168–176.

Chapman, C. S., Gallivan, J. P., Wood, D. K., Milne, J. L., Culham, J. C., & Goodale, M. A. (2010b). Short-term motor plasticity revealed in a visuomotor decision-making task. *Behavioural Brain Research*, *214*(1), 130–134.

Cheng, K. Y., Chapman, C. S., & Hebert, J. S. (2022). Spatiotemporal coupling of hand and eye movements when using a myoelectric prosthetic hand [ISSN: 1945-7901]. *2022 International Conference on Rehabilitation Robotics (ICORR)*, 1–6.

Cheung, J. H., Burns, D. K., Sinclair, R. R., & Sliter, M. (2017). Amazon mechanical turk in organizational psychology: An evaluation and practical recommendations. *Journal of Business and Psychology*, *32*(4), 347–361.

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? shifts in data quality and the impact on study results [Publisher: SAGE Publications Inc]. *Social Psychological and Personality Science*, *11*(4), 464–473.

Cisek, P. (2022). Evolution of behavioural control from chordates to primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1844), 20200522.

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, *33*(1), 269–298.

Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? an experimental comparison of data quality in laboratory and online studies [Publisher: Cambridge University Press]. *Journal of Experimental Political Science*, *1*(2), 120–131.

Cohen, J., Collins, R., Darkes, J., & Gwartney, D. (2007). A league of their own: Demographics, motivations and patterns of use of 1,955 male adult non-medical anabolic steroid users in the united states [Publisher: Routledge _eprint: https://doi.org/10.1186/1550-2783-4-12]. *Journal of the International Society of Sports Nutrition*, *4*(1), 12.

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should i stay or should i go? how the human brain manages the trade-off between exploitation and exploration [Publisher: Royal Society]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933–942.

Cramer, A. O. J., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R. P. P. P., Waldorp, L. J., & Wagenmakers, E.-J. (2016). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic Bulletin & Review*, *23*(2), 640–647.

Cristina, S., & Camilleri, K. P. (2018). Unobtrusive and pervasive video-based eye-gaze tracking. *Image and Vision Computing*, *74*, 21–40.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research (S. Gilbert, Ed.). *PLoS ONE*, *8*(3), e57410.

Cuttler, C., LaFrance, E. M., & Stueber, A. (2021). Acute effects of high-potency cannabis flower and cannabis concentrates on everyday life memory and decision making [Number: 1 Publisher: Nature Publishing Group]. *Scientific Reports*, *11*(1), 13784.

Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science*, *35*(5), 983–996.

Danion, F. R., & Flanagan, J. R. (2018). Different gaze strategies during eye versus hand tracking of a moving target. *Scientific Reports*, *8*(1), 10059.

Deconinck, F. J. A., van Polanen, V., Savelsbergh, G. J. P., & Bennett, S. J. (2011). The relative timing between eye and hand in rapid sequential pointing is affected by time pressure, but not by advance knowledge. *Experimental Brain Research*, *213*(1), 99–109.

Delabarre, E. B. (1898). A method of recording eye-movements [Publisher: University of Illinois Press]. *The American Journal of Psychology*, *9*(4), 572–574.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12.

Deng, S., Chang, J., Kirkby, J. A., & Zhang, J. J. (2016). Gaze–mouse coordinated movements and dependency with coordination demands in tracing. *Behaviour & Information Technology*, *35*(8), 665–679.

DeVoe, S. E., & House, J. (2016). Replications with MTurkers who are naïve versus experienced with academic studies: A comment on connors, khamitov, moroz, campbell, and henderson (2015). *Journal of Experimental Social Psychology*, *67*, 65–67.

Donegan, M., Morris, J. D., Corno, F., Signorile, I., Chió, A., Pasian, V., Vignola, A., Buchholz, M., & Holmqvist, E. (2009). Understanding users and their needs. *Universal Access in the Information Society*, *8*(4), 259–275.

Dotan, D., Meyniel, F., & Dehaene, S. (2018). On-line confidence monitoring during decision making. *Cognition*, *171*, 112–121.
Dotan, D., Pinheiro-Chagas, P., Al Roumi, F., & Dehaene, S. (2019). Track it to crack it: Dissecting processing stages with finger tracking. *Trends in Cognitive Sciences*, *23*(12), 1058–1070.
Duewer, D. L. (2022). *Face recognition vendor test (FRVT) part 8:: Summarizing demographic differentials* (NIST IR 8429). National Institute of Standards and Technology. Gaithersburg, MD.
Dupuis, M. (2019). Detecting computer-generated random responding in questionnaire-based data: A comparison of seven indices, 10.
Ehinger, B. V., Groß, K., Ibs, I., & König, P. (2019). A new comprehensive eye-tracking test battery concurrently evaluating the pupil labs glasses and the EyeLink 1000 [Publisher: PeerJ Inc.]. *PeerJ*, *7*, e7086.
EyeLink 1000 plus technical specifications. (2017). SR Research. Retrieved July 18, 2023, from https://www.sr-research.com/wp-content/uploads/2017/11/eyelink-1000-plus-specifications.pdf
Faulkenberry, T. J., Cruise, A., Lavro, D., & Shaki, S. (2016). Response trajectories capture the continuous dynamics of the size congruity effect. *Acta Psychologica*, *163*, 114–123.
Federico, G., & Brandimonte, M. A. (2019). Tool and object affordances: An ecological eye-tracking study. *Brain and Cognition*, *135*, 103582.
Federico, G., Ferrante, D., Marcatto, F., & Brandimonte, M. A. (2021). How the fear of COVID-19 changed the way we look at human faces [Publisher: PeerJ Inc.]. *PeerJ*, *9*, e11380.
Federico, G., Osiurak, F., Brandimonte, M. A., Salvatore, M., & Cavaliere, C. (2023). The visual encoding of graspable unfamiliar objects. *Psychological Research*, *87*(2), 452–461.
Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: A unified JavaScript framework for online studies. *International Conference on Computational Social Science*, *1*(1), 1–3.
Fitts, P. M. (n.d.). REPRINT VERSION the information capacity of the human motor system in controlling the amplitude of movement, 8.
Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 11.
Fraser, A., Gattas, S., Hurman, K., Robinson, M., Duta, M., & Scerif, G. (2021, June 22). Automated gaze direction scoring from videos collected online through conventional webcam.
Freeman, J. B. (2018). Doing psychological science by hand [Publisher: SAGE Publications Inc]. *Current Directions in Psychological Science*, *27*(5), 315–323.
Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1436), 2435–2440. Retrieved September 13, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1690464/

Gagné, N., & Franzen, L. (2023). How to run behavioural experiments online: Best practice suggestions for cognitive psychology and neuroscience. *Swiss Psychology Open: the official journal of the Swiss Psychological Society*, *3*(1), 1.

Gallivan, J. P., Chapman, C. S., Wolpert, D. M., & Flanagan, J. R. (2018). Decision-making in sensorimotor control [Number: 9 Publisher: Nature Publishing Group]. *Nature Reviews Neuroscience*, *19*(9), 519–534.

Garaizar, P., & Reips, U.-D. (2019). Best practices: Two web-browser-based methods for stimulus presentation in behavioral experiments with high-resolution timing requirements. *Behavior Research Methods*, *51*(3), 1441–1453.

Garvie, C., Bedoya, A. M., Frankle, J., Daugherty, M., Evans, K., George, E. J., McCubbin, S., Rudolph, H., Ullman, I., Ainsworth, S., Houck, D., Iorio, M., Kahn, M., Olson, E., Petenko, J., & Singleton, K. (2016). UNREGULATED POLICE FACE RECOGNITION IN AMERICA.

Glaholt, M. G., & Reingold, E. M. (2009). The time course of gaze bias in visual decision tasks. *Visual Cognition*, *17*(8), 1228–1243.

Glaholt, M. G., Wu, M.-C., & Reingold, E. M. (2009). Predicting preference from fixations [Place: Italy Publisher: PsychNology Journal]. *PsychNology Journal*, *7*, 141–158.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making [_eprint: https://doi.org/10.1146/annurev.neuro.29.051605.113038]. *Annual Review of Neuroscience*, *30*(1), 535–574.

Gottlieb, J. (2018). Understanding active sampling strategies: Empirical approaches and implications for attention and decision research. *Cortex*, *102*, 150–160.

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information seeking, curiosity and attention: Computational and neural mechanisms. *Trends in cognitive sciences*, *17*(11), 585–593.

Greenaway, A.-M., Nasuto, S., Ho, A., & Hwang, F. (2021). Is home-based webcam eye-tracking with older adults living with and without alzheimer's disease feasible? *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, 1–3.

Grootswagers, T. (2020). A primer on running human behavioural experiments online. *Behavior Research Methods*, *52*(6), 2283–2286.

Grossman, T., & Balakrishnan, R. (2004). Pointing at trivariate targets in 3d environments. *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, 447–454.

Gudi, A., Li, X., & van Gemert, J. (2020). Efficiency in real-time webcam gaze tracking. In A. Bartoli & A. Fusiello (Eds.), *Computer vision – ECCV 2020 workshops* (pp. 529–543). Springer International Publishing.

Guo, Q., & Agichtein, E. (2010a). Ready to buy or just browsing?: Detecting web searcher goals from interaction data. *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10*, 1–10.

Guo, Q., & Agichtein, E. (2010b). Towards predicting web searcher gaze position from mouse movements. *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, 3601–3606.

Hacker, P., Engel, A., & Mauer, M. (2023). Regulating ChatGPT and other large generative AI models. *2023 ACM Conference on Fairness, Accountability, and Transparency*, 1112–1123.

Hansen, D. W., & Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze [Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 478–500.

Harrison, P. M. c. (2020). psychTestR: An r package for designing and conducting behavioural psychological experiments. *Journal of Open Source Software*, *5*(49), 2088.

Hartshorne, J. K., de Leeuw, J. R., Goodman, N. D., Jennings, M., & O'Donnell, T. J. (2019). A thousand studies for the price of one: Accelerating psychological science with pushkin. *Behavior Research Methods*, *51*(4), 1782–1803.

Hauger, D., Paramythis, A., & Weibelzahl, S. (2011). Using browser interaction data to determine page reading behavior. In J. A. Konstan, R. Conejo, J. L. Marzo, & N. Oliver (Eds.), *User modeling, adaption and personalization* (pp. 147–158). Springer.

Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, *7*(1), 43–64.

Hebert, J. S., Boser, Q. A., Valevicius, A. M., Tanikawa, H., Lavoie, E. B., Vette, A. H., Pilarski, P. M., & Chapman, C. S. (2019). Quantitative eye gaze and movement differences in visuomotor adaptations to varying task demands among upper-extremity prosthesis users. *JAMA Network Open*, *2*(9), e1911197.

Heck, M., Becker, C., & Deutscher, V. (2023, January 3). *Webcam eye tracking for desktop and mobile devices: A systematic review*. Retrieved July 13, 2023, from https://hdl.handle.net/10125/103459

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, *18*(3), 384–401.

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). Lab.js: A free, open, online study builder. *Behavior Research Methods*, *54*(2), 556–573.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD [Number: 7302 Publisher: Nature Publishing Group]. *Nature*, *466*(7302), 29–29.

Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The 'real-world approach' and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, *11*, 721.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. v. d. (2011, September 22). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Hsu, K. J., Caffey, K., Pisner, D., Shumake, J., Risom, S., Ray, K. L., Smits, J. A. J., Schnyer, D. M., & Beevers, C. G. (2018). Attentional bias modification treatment for depression: Study protocol for a randomized controlled trial. *Contemporary Clinical Trials*, *75*, 59–66.

Huang, J., White, R., & Buscher, G. (2012). User see, user point: Gaze and cursor alignment in web search. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1341–1350.

Hutt, S., Wong, A., Papoutsaki, A., Baker, R. S., Gold, J. I., & Mills, C. (2023). Webcam-based eye tracking to detect mind wandering and comprehension errors. *Behavior Research Methods*.

IAPP privacy and consumer trust report. (2023, March). International Association of Privacy Professionals. Retrieved August 17, 2023, from https://iapp.org/media/pdf/resource_center/privacy_and_consumer_trust_report_summary.pdf

Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, cambridge analytica, and privacy protection [Conference Name: Computer]. *Computer*, *51*(8), 56–59.

Jacobs, O. L. E., Pazhoohi, F., & Kingstone, A. (2023). Contrapposto posture captures visual attention: An online gaze tracking experiment [Publisher: Routledge _eprint: https://doi.org/10.1080/13506285.2023.2213904]. *Visual Cognition*, *31*(2), 160–167.

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye–hand coordination in object manipulation. *The Journal of Neuroscience*, *21*(17), 6917–6932.

Johnson, B. P., Dayan, E., Censor, N., & Cohen, L. G. (2021). Crowdsourcing in cognitive and systems neuroscience [Publisher: SAGE Publications Inc STM]. *The Neuroscientist*, 10738584211017018.

Jun, E., Hsieh, G., & Reinecke, K. (2017). Types of motivation affect study selection, attention, and dropouts in online experiments. *Proceedings of the ACM on Human-Computer Interaction*, *1*, 56:1–56:15.

Kar, A., & Corcoran, P. (2016). Towards the development of a standardized performance evaluation framework for eye gaze estimation systems in consumer platforms. *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 002061–002066.

Karousos, N., Katsanos, C., Tselios, N., & Xenos, M. (2013). Effortless tool-based evaluation of web form filling tasks using keystroke level model and fitts law. *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, 1851.

Kerr, R. (1973). Movement time in an underwater environment. *Journal of Motor Behavior*, *5*(3), 175–178.

Khamis, M., Alt, F., & Bulling, A. (2018). The past, present, and future of gaze-enabled handheld mobile devices: Survey and lessons learned. *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–17.

Khamis, M., Alt, F., Hassib, M., von Zezschwitz, E., Hasholzner, R., & Bulling, A. (2016). GazeTouchPass: Multimodal authentication using gaze and touch on mobile devices. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2156–2164.

Khamis, M., Hassib, M., Zezschwitz, E. v., Bulling, A., & Alt, F. (2017). GazeTouch-PIN: Protecting sensitive data on mobile devices using secure multimodal authentication. *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 446–450.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317–340.

Klaib, A. F., Alsrehin, N. O., Melhem, W. Y., & Bashtawi, H. O. (2019). IoT smart home using eye tracking and voice interfaces for elderly and special needs people. *Journal of Communications*, 614–621.

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, *67*(4), 151–185.

Krajbich, I. (2019). Accounting for attention in sequential sampling models of decision making. *Current Opinion in Psychology*, *29*, 6–11.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology*, *3*. Retrieved May 25, 2023, from https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00193

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13852–13857.

Kröger, J. L., Lutz, O. H.-M., & Müller, F. (2020). What does your gaze reveal about you? on the privacy implications of eye tracking. In M. Friedewald, M. Önen, E. Lievens, S. Krenn, & S. Fricker (Eds.), *Privacy and identity management. data for better living: AI and privacy: 14th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2.2 international summer school, windisch, switzerland, august 19–23, 2019, revised selected papers* (pp. 226–241). Springer International Publishing.

Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, *25*(3), 296–324.

Land, M. F. (2009). Vision, eye movements, and natural behavior. [MAG ID: 2120779958]. *Visual Neuroscience*, *26*(1), 51–62.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities. *Vision Research*, *41*(25), 3559–3565.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328.

Lang, P., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS) : Affective ratings of pictures and instruction manual [Publisher: University of Florida]. *Technical Report*. Retrieved May 31, 2023, from https://cir.nii.ac.jp/crid/1573950399053852928

Lariviere, J. A. (2015). Eye tracking: Eye-gaze technology. In I. Söderback (Ed.), *International handbook of occupational therapy interventions* (pp. 339–362). Springer International Publishing.
Lavoie, E. B., Valevicius, A. M., Boser, Q. A., Kovic, O., Vette, A. H., Pilarski, P. M., Hebert, J. S., & Chapman, C. S. (2018). Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. *Journal of Vision*, *18*(6), 1–20.
Le, T., Dietz, F., Pfeuffer, K., & Alt, F. (2022). A practical method to eye-tracking on the phone: Toolkit, accuracy and precision. *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia*, 182–188.
Leeuw, J. R. d., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351.
Levine, J. L. (1981). *An eye-controlled computer*. IBM Research Division, TJ Watson Research Center.
Liebling, D., & Dumais, S. (2014). Gaze and mouse coordination in everyday work. *UBICOMP ADJUNCT '14*, 10.
Lin, C. J., & Ho, S.-H. (2020). Prediction of the use of mobile device interfaces in the progressive aging process with the model of fitts' law. *Journal of Biomedical Informatics*, *107*, 103457.
Logitech 2023 annual report, SEC form 10-k (Annual Report). (2023). Logitech. Retrieved August 24, 2023, from https://s1.q4cdn.com/104539020/files/doc_financials/2023/ar/71e634f7-ce7a-4c9c-95cf-5a98d8fb9068.pdf
Lourenco, S. F., & Tasimi, A. (2020). No participant left behind: Conducting science during COVID-19. *Trends in Cognitive Sciences*, *24*(8), 583–584.
Lovett, M., Bajaba, S., Lovett, M., & Simmering, M. J. (2018). Data quality from crowdsourced surveys: A mixed method inquiry into perceptions of amazon's mechanical turk masters [Publisher: John Wiley & Sons, Ltd]. *Applied Psychology*, *67*(2), 339–366.
Lukács, G., & Gartus, A. (2022). Precise display time measurement in JavaScript for web-based experiments. *Behavior Research Methods*.
Ma, H. L., Bertrand, J. K., Chapman, C. S., & Hayward, D. A. (2023). You read my mind: Generating and minimizing intention uncertainty under different social contexts in a two-player online game [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: Human Perception and Performance*, No Pagination Specified–No Pagination Specified.
MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human–Computer Interaction*, *7*(1), 91–139.
MacKenzie, I. S., & Buxton, W. (1992). Extending fitts' law to two-dimensional tasks. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92*, 219–226.
Mackenzie, I. S., Zhang, S. X., & Soukoreff, R. W. (1999). Text entry using soft keyboards. *Behaviour & Information Technology*, *18*(4), 235–244.
Madsen, J., Júlio, S. U., Gucik, P. J., Steinberg, R., & Parra, L. C. (2021). Synchronized eye movements predict test scores in online video education [Publisher:

Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *118*(5), e2016980118.

Maldonado, M., Dunbar, E., & Chemla, E. (2019). Mouse tracking as a window into decision making. *Behavior Research Methods*, *51*(3), 1085–1101.

Mason, W., & Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavior Research Methods*, *44*(1), 1–23.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of p300 latency and reaction time [Publisher: American Association for the Advancement of Science]. *Science*, *211*(4477), 77–80.

McGuffin, M. J., & Balakrishnan, R. (2005). Fitts' law and expanding targets: Experimental studies and designs for user interfaces. *ACM Transactions on Computer-Human Interaction*, *12*(4), 35.

Meng, C., & Zhao, X. (2017). Webcam-based eye movement analysis using CNN [Conference Name: IEEE Access]. *IEEE Access*, *5*, 19581–19587.

Miller, R., Schmidt, K., Kirschbaum, C., & Enge, S. (2018). Comparability, stability, and reliability of internet-based mental chronometry in domestic and laboratory settings. *Behavior Research Methods*, *50*(4), 1345–1358.

Moher, J., & Song, J.-H. (2014). Perceptual decision processes flexibly adapt to avoid change-of-mind motor costs. *Journal of Vision*, *14*(8), 1.

Moss, A., Rosenzweig, C., Jaffe, S. N., Gautam, R., Robinson, J., & Litman, L. (2021). *Bots or inattentive humans? identifying sources of low-quality data in online platform* (Type: article).

Moss, A. J., Rosenzweig, C., Robinson, J., Jaffe, S. N., & Litman, L. (2023). Is it ethical to use mechanical turk for behavioral research? relevant data from a representative survey of MTurk participants and wages. *Behavior Research Methods*.

Namnakani, O., Abdrabou, Y., Grizou, J., Esteves, A., & Khamis, M. (2023). Comparing dwell time, pursuits and gaze gestures for gaze interaction on handheld mobile devices. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–17.

Navalpakkam, V., Jentzsch, L., Sayres, R., Ravi, S., Ahmed, A., & Smola, A. (2013). Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. *Proceedings of the 22nd international conference on World Wide Web - WWW '13*, 953–964.

Necka, E. A., Cacioppo, S., Norman, G. J., & Cacioppo, J. T. (2016). Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community participants [Publisher: Public Library of Science]. *PLOS ONE*, *11*(6), e0157732.

Newman, A., Bavik, Y. L., Mount, M., & Shao, B. (2021). Data collection via online platforms: Challenges and recommendations for future research. *Applied Psychology*, *70*(3), 1380–1402.

Newman, D., & Lathan, C. (1999). Memory processes and motor control in extreme environments. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *29*(3), 387–394.

Ngo, V., Gorman, J. C., De la Fuente, M. F., Souto, A., Schiel, N., & Miller, C. T. (2022). Active vision during prey capture in wild marmoset monkeys. *Current Biology*, *32*(15), 3423–3428.e3.

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206.

Ouellette Zuk, A. A., Bertrand, J. K., & Chapman, C. S. (2023, June 7). Continuous measures of decision-difficulty captured remotely: I. mouse-tracking sensitivity extends to tablets and smartphones [Pages: 2023.06.06.543796 Section: New Results].

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*(5), 1.

Panait, L., Akkary, E., Bell, R. L., Roberts, K. E., Dudrick, S. J., & Duffy, A. J. (2009). The role of haptic feedback in laparoscopic simulation training. *Journal of Surgical Research*, *156*(2), 312–316.

Paolacci, G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, *5*(5), 9.

Papoutsaki, A., Laskey, J., & Huang, J. (2017). SearchGazer: Webcam eye tracking for remote studies of web search. *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, 17–26.

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions.

Patla, A., & Vickers, J. (2003). How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental Brain Research*, *148*(1), 133–138.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Pelz, J., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, *139*(3), 266–277.

Policy statement of the FTC on biometric information and section 5 of the FTC act (Policy Statement). (2023). Federal Trade Commission. Retrieved August 25, 2023, from https://www.ftc.gov/system/files/ftc_gov/pdf/p225402biometricpolicystatement.pdf

Pronk, T., Wiers, R. W., Molenkamp, B., & Murre, J. (2020). Mental chronometry in the pocket? timing accuracy of web applications on touchscreen and keyboard devices. *Behavior Research Methods*, *52*(3), 1371–1382.

Rand, M. K., & Stelmach, G. E. (2010). Effects of hand termination and accuracy constraint on eye–hand coordination during sequential two-segment movements. *Experimental Brain Research*, *207*(3), 197–211.

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology*, *20*(2), 262–270.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions [Place: United Kingdom Publisher: Blackwell Publishing]. *Psychological Science*, *9*, 347–356.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search [Publisher: SAGE Publications]. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

Rayner, K., & Reingold, E. M. (2015). Evidence for direct cognitive control of fixation durations during reading. *Current Opinion in Behavioral Sciences*, *1*, 107–112.

Ringrose, K. (2019). Law enforcement's pairing of facial recognition technology with body-worn cameras escalates privacy concerns. *Virginia Law Review Online*, *105*, 57. https://heinonline.org/HOL/Page?handle=hein.journals/inbrf105&id=57&div=&collection=

Robal, T., Zhao, Y., Lofi, C., & Hauff, C. (2018). Webcam-based attention tracking in online learning: A feasibility study. *23rd International Conference on Intelligent User Interfaces*, 189–197.

Rodden, K., & Fu, X. (2007). Exploring how mouse movements relate to eye movements on web search results pages. *Workshop on Web Information Seeking and Interaction at SIGIR '07*, 29–32. Retrieved September 12, 2022, from https://storage.googleapis.com/pub-tools-public-publication-data/pdf/32735.pdf#page=33

Roy, N., Câmara, A., Maxwell, D., & Hauff, C. (2021). Incorporating widget positioning in interaction models of search behaviour. *roceedings of the 2021 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR '21)*, 10.

Sailer, U., Eggert, T., Ditterich, J., & Straube, A. (2000). Spatial and temporal aspects of eye-hand coordination across different tasks. *Experimental Brain Research*, *134*, 163–173.

Salminen, J., Jansen, B. J., An, J., Jung, S.-G., Nielsen, L., & Kwak, H. (2018). Fixation and confusion: Investigating eye-tracking participants' exposure to information in personas. *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, 110–119.

San Agustin, J., Skovsgaard, H., Mollenbach, E., Barret, M., Tall, M., Hansen, D. W., & Hansen, J. P. (2010). Evaluation of a low-cost open-source gaze tracker. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, 77–80.

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences*, *10*(4), 251.

Saxena, S., Lange, E., & Fink, L. (2022). Towards efficient calibration for webcam eye-tracking in online experiments. *2022 Symposium on Eye Tracking Research and Applications*, 1–7.

Schneegans, T., Bachman, M., Huettel, S., & Heekeren, H. (2021). *Exploring the potential of online webcam-based eye tracking in decision-making research and influence factors on data qualit* (Type: article).

Schouten, J. F., & Bekker, J. A. M. (1967). Reaction time and accuracy. *Acta Psychologica*, *27*, 143–153.

Schröter, I., Grillo, N. R., Limpak, M. K., Mestiri, B., Osthold, B., Sebti, F., & Mergenthaler, M. (2021). Webcam eye tracking for monitoring visual attention in hypothetical online shopping tasks [Number: 19 Publisher: Multidisciplinary Digital Publishing Institute]. *Applied Sciences*, *11*(19), 9281.

Schuetz, I., Murdison, T. S., MacKenzie, K. J., & Zannoli, M. (2019). An explanation of fitts' law-like performance in gaze-based selection tasks using a psychophysics approach. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13.

Scott, K., & Schulz, L. (2017). Lookit (part 1): A new online platform for developmental research. *Open Mind*, *1*(1), 4–14.

Semmelmann, K., Hönekopp, A., & Weigelt, S. (2017). Looking tasks online: Utilizing webcams to collect video data from home. *Frontiers in Psychology*, *8*(1582), 1–11.

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451–465.

Seow, S. C. (2005). Information theoretic models of HCI: A comparison of the hick-hyman law and fitts' law. *Human–Computer Interaction*, *20*(3), 315–352.

Sewell, W., & Komogortsev, O. (2010). Real-time eye gaze tracking with an unmodified commodity webcam employing a neural network. *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, 3739–3744.

Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-life neuroscience: An ecological approach to brain and behavior research [Publisher: SAGE Publications Inc]. *Perspectives on Psychological Science*, *14*(5), 841–859.

Shehu, I. S., Wang, Y., Athuman, A. M., & Fu, X. (2021a). Paradigm shift in remote eye gaze tracking research: Highlights on past and recent progress. In K. Arai, S. Kapoor, & R. Bhatia (Eds.), *Proceedings of the future technologies conference (FTC) 2020, volume 1* (pp. 159–189). Springer International Publishing.

Shehu, I. S., Wang, Y., Athuman, A. M., & Fu, X. (2021b). Remote eye gaze tracking research: A comparative evaluation on past and recent progress [Number: 24 Publisher: Multidisciplinary Digital Publishing Institute]. *Electronics*, *10*(24), 3165.

Shevchenko, Y. (2022). Open lab: A web application for running and sharing online experiments. *Behavior Research Methods*, *54*(6), 3118–3125.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317–1322.

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2011). Adaptive allocation of vision under competing task demands. *Journal of Neuroscience*, *31*(3), 928–943.

Skovsgaard, H., Agustin, J. S., Johansen, S. A., Hansen, J. P., & Tall, M. (2011). Evaluation of a remote webcam-based eye tracker. *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications*, 1–4.

Slim, M. S., & Hartsuiker, R. J. (2022). Moving visual world experiments online? a web-based replication of dijkgraaf, hartsuiker, and duyck (2017) using PCIbex and WebGazer.js. *Behavior Research Methods*.

Smith, B. A., Ho, J., Ark, W., & Zhai, S. (2000). Hand eye coordination patterns in target selection. *Eye Tracking Research & Applications Symposium*, 117–122.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*(2), 135–168.

Smith, S. M., & Krajbich, I. (2018). Attention and choice across domains. *Journal of Experimental Psychology: General*, *147*(12), 1810–1826.

Stewart, E., Valsecchi, M., & Schütz, A. C. (2020). A review of interactions between peripheral and foveal vision. *Journal of Vision*, *20*(12), 2.

Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences*, *21*(10), 736–748.

Stillman, P. E., Krajbich, I., & Ferguson, M. J. (2020). Using dynamic monitoring of choices to predict and understand risk preferences [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *117*(50), 31738–31747.

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences*, *22*(6), 531–543.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments [Publisher: SAGE Publications Inc]. *Teaching of Psychology*, *44*(1), 24–31.

Stone, S. A., Boser, Q. A., Dawson, T. R., Vette, A. H., Hebert, J. S., Pilarski, P. M., & Chapman, C. S. (2022). Generating accurate 3d gaze vectors using synchronized eye tracking and motion capture. *Behavior Research Methods*.

Stone, S. A., & Chapman, C. S. (2023). Unconscious frustration: Dynamically assessing user experience using eye and mouse tracking. *Proceedings of the ACM on Human-Computer Interaction*, *7*, 168:1–168:17.

Thomas, K. A., & Clifford, S. (2017). Validity and mechanical turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, *77*, 184–197.

Uittenhove, K., Jeanneret, S., & Vergauwe, E. (2022). *From lab-testing to web-testing in cognitive research: Who you test is more important than how you tes* (Type: article).

Valevicius, A. M., Boser, Q. A., Lavoie, E. B., Murgatroyd, G. S., Pilarski, P. M., Chapman, C. S., Vette, A. H., & Hebert, J. S. (2018). Characterization of normative hand movements during two functional upper limb tasks. *PLOS ONE*, *13*(6), e0199549.

Valliappan, N., Dai, N., Steinberg, E., He, J., Rogers, K., Ramachandran, V., Xu, P., Shojaeizadeh, M., Guo, L., Kohlhoff, K., & Navalpakkam, V. (2020). Accelerating eye movement research via accurate and affordable smartphone eye tracking. *Nature Communications*, *11*(1), 4553.

Van Doorn, G., Woods, A., Levitan, C. A., Wan, X., Velasco, C., Bernal-Torres, C., & Spence, C. (2017). Does the shape of a cup influence coffee taste expectations? a cross-cultural, online study. *Food Quality and Preference*, *56*, 201–211.

van Donkelaar, P., & Staub, J. (2000). Eye-hand coordination to visual versus remembered targets. *Experimental Brain Research*, *133*(3), 414–418.

Veselovsky, V., Ribeiro, M. H., & West, R. (2023, June 13). Artificial artificial artificial intelligence: Crowd workers widely use large language models for text production tasks. Retrieved July 27, 2023, from http://arxiv.org/abs/2306.07899

Vos, M., Minor, S., & Ramchand, G. C. (2022). Comparing infrared and webcam eye tracking in the visual world paradigm [Accepted: 2023-01-03T13:11:04Z Publisher: University of California Press].

Wadsworth, H. E., Galusha-Glasscock, J. M., Womack, K. B., Quiceno, M., Weiner, M. F., Hynan, L. S., Shore, J., & Cullum, C. M. (2016). Remote neuropsychological assessment in rural american indians with and without cognitive impairment. *Archives of Clinical Neuropsychology*, *31*(5), 420–425.

Walter, S. L., Seibert, S. E., Goering, D., & O'Boyle, E. H. (2019). A tale of two sample sources: Do results from online panel data and conventional data converge? *Journal of Business and Psychology*, *34*(4), 425–452.

Weydmann, G., Palmieri, I., Simões, R. A. G., Centurion Cabral, J. C., Eckhardt, J., Tavares, P., Moro, C., Alves, P., Buchmann, S., Schmidt, E., Friedman, R., & Bizarro, L. (2022). Switching to online: Testing the validity of supervised remote testing for online reinforcement learning experiments. *Behavior Research Methods*.

William Soukoreff, R., & Scott Mackenzie, I. (1995). Theoretical upper and lower bounds on typing speed using a stylus and a soft keyboard. *Behaviour & Information Technology*, *14*(6), 370–379.

Williams, H. E., Boser, Q. A., Pilarski, P. M., Chapman, C. S., Vette, A. H., & Hebert, J. S. (2019). Hand function kinematics when using a simulated myoelectric prosthesis [ISSN: 1945-7901]. *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*, 169–174.

Wisiecka, K., Krejtz, K., Krejtz, I., Sromek, D., Cellary, A., Lewandowska, B., & Duchowski, A. (2022). Comparison of webcam and remote eye tracking. *2022 Symposium on Eye Tracking Research and Applications*, 1–7.

Wispinski, N. J., Gallivan, J. P., & Chapman, C. S. (2020). Models, movements, and minds: Bridging the gap between decision making and action. *Annals of the New York Academy of Sciences*, *1464*(1), 30–51.

Wong, A. Y., Bryck, R. L., Baker, R. S., Hutt, S., & Mills, C. (2023). Using a webcam based eye-tracker to understand students' thought patterns and reading behaviors in neurodivergent classrooms. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 453–463.

Xu, P., Ehinger, K. A., Zhang, Y., Finkelstein, A., Kulkarni, S. R., & Xiao, J. (2015, May 20). TurkerGaze: Crowdsourcing saliency with webcam based eye tracking. Retrieved August 24, 2022, from http://arxiv.org/abs/1504.06755

Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, *16*(6), 1485–1505. Retrieved August 24, 2022, from https://journal.sjdm.org/21/210525/jdm210525.html

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In A. L. Yarbus (Ed.), *Eye movements and vision* (pp. 171–211). Springer US.

Young, L. R., & Sheena, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, *7*(5), 397–429.

Zehr, J., & Schwarz, F. (2018). PennController for internet based experiments (IBEX) [Publisher: OSF].

Zhang, X., Sugano, Y., & Bulling, A. (2019). Evaluation of appearance-based methods and implications for gaze-based applications. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13.

Zhao, Y., Lofi, C., & Hauff, C. (2017). Scalable mind-wandering detection for MOOCs: A webcam-based approach. In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, & M. Pérez-Sanagustín (Eds.), *Data driven approaches in digital education* (pp. 330–344). Springer International Publishing.

Zheng, C., & Usagawa, T. (2018). A rapid webcam-based eye tracking method for human computer interaction [ISSN: 2475-7896]. *2018 International Conference on Control, Automation and Information Sciences (ICCAIS)*, 133–136.

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, *111*(4), 493–504.

Appendix A: Continuous Measures of Decision-Difficulty Captured Remotely: I. Mouse-tracking sensitivity extends to tablets and smartphones (Ouellette Zuk, Bertrand & Chapman, 2023)

Continuous Measures of Decision-Difficulty Captured Remotely: I. Mouse-tracking sensitivity extends to tablets and smartphones

Alexandra A. Ouellette Zuk[1] Jennifer K. Bertrand[1,2] Craig S. Chapman[1,2]

[1]Neuroscience and Mental Health Institute, University of Alberta
[2]Faculty of Kinesiology, Sport, and Recreation, University of Alberta

Abstract

As decisions require actions to have an effect on the world, measures derived from movements such as using a mouse to control a cursor on a screen provide powerful and dynamic indices of decision-making. In this first of a set of two studies, we replicated classic reach-decision paradigms across computers, tablets, and smartphones, we show that portable touch-devices can sensitively capture decision-difficulty. We see this in pre- and during-movement temporal and motoric measures across diverse decision domains. We found touchscreen interactions to more sensitively reflect decision-difficulty during movement compared to computer interactions, and the latter to be more sensitive before movement initiation. Paired with additional evidence for the flexibility and unique utility of pre- and during-movement measures, this substantiates the use of widely available touch-devices to massively extend the reach of decision science. We build upon this in the second study in this series (Bertrand et al., 2023) with the use of webcam eye-tracking to further elucidate, earlier in time, the decision process. This subsequent work provides additional support for tools that enable remote collection of rich decision data in ecologically-valid environments.

Keywords: decision-making, mouse-tracking, tablets, smartphones, remote data capture

1 Introduction

Our lives unfold as an amalgamation of decisions made and actions taken to execute them. Ultimately, these enacted choices shape our lives and our societies. As a result, the study of human decision behaviour has inspired researchers for centuries, from interest in risk preference amongst gamblers [5], to willingness to pay given prior value contexts [27].

Historically, most measures of decision-making use verbal reports (e.g., [38, 27]), observed choices (e.g., [34]), or discrete measurements of behaviour such as reaction time and accuracy (see [41] for review). Reaction times, specifically, have been shown to reflect cognitive conflict during decision-making, with more difficult decisions leading to longer reaction times [32, 36, 40]. These approaches, which focus almost exclusively on the outcome of a decision, fail to account for the embodied nature of real-world decision-making. In the real-world, a decision is not made until a body physically enacts the choice. Recognizing that *how* we decide is likely as important as *what* we decide, researchers have started recording the dynamics of behaviour [9, 22, 14, 15, 50]. Requiring and tracking movement to select between choices, reach-decision paradigms are a popular method for continuously measuring the factors that underlie and bias the decision process. These tasks have quantified decision behaviours across a variety of choice domains for both real 3-D reaching [8, 7, 21, 22] and for 2-D computer-mouse tracking [20, 44, 26, 45].

Computerized reach-decision tasks, with 2-D movements made by a computer-mouse are a particularly sensitive, flexible, and scalable technique for the examination of decision processes ([28, 33, 26, 17, 20, 44, 31] and many more). Requiring participants to start with their mouse cursor centered at the bottom of the computer screen and necessitating the selection of one of two (most commonly) choice options located in the top left or right corners of the screen, classic mouse-tracking paradigms record the attraction toward each of the two choice options. This generates a vertical movement component relatively independent of the competition between options (though, movement speed has been related to different aspects of the decision process [14, 15]) and a critical horizontal movement component that tracks either directly toward one of the two options when there is no choice-competition, or indirectly between the two options when the choice-competition is high [14, 15, 44]. The typical result is a continuum of direct to indirect trajectories, reflecting the strength of competition between choice options and thus the relative difficulty of the decision. Metrics quantifying relative reach directness include the maximum absolute deviation from a straight trajectory and movement times. Like pre-movement reaction times, these during-movement measures of movement time and curvature are also sensitive to decision-difficulty, with harder decisions resulting in longer duration movements and greater trajectory curvature (as seen in Figure 1 and [28, 26, 17, 20, 44, 31, 45]).

Despite reach-decision trajectory-tracking being an important tool for the understanding of decision-making, these approaches remain relatively unused outside of research labs. Recognizing that research deployed online via portable devices could reach a wider and more diverse audience, there has been a recent movement to assess the reliability of cognitive task administration in these environments [2, 39, 37]. This has been fuelled by new tools allowing the development of online tasks (e.g., Labvanced [18], Gorilla [3], jsPsych [30]) that include easy deployment to diverse, crowd-sourced participant pools (e.g MTurk [1], Prolific [35]) and can target a variety of devices [2].

While cognitive tasks measuring accuracy and reaction time have been replicated on tablets [19, 42] and smartphones [4], it is largely unknown if and how motoric measures of decision-difficulty can be measured on these portable devices. To test this question, we developed a reach-decision task using Labvanced [18] to collect continuous cursor position data, and deployed it to over 300 crowd-sourced participants. Critically, each of these participants completed the task on one of three different devices (>100 participants per device) varying in size and user-interaction requirements: personal computers (mouse-based interactions), tablets (finger or stylus-based interactions) and smartphones (finger-,thumb- or stylus-based interactions).

To provide evidence that a particular device is tracking decision-difficulty, we chose to replicate three unique reach-decision tasks. Each of these tasks has been shown to sensitively reflect decision-difficulty effects through mouse-tracking (see Figure 1A) and here we tested if those effects were replicable and then extensible to tablets and smartphones. The three tasks were: a Numeric-Size Congruity task [17], a Sentence Verification task [13] and a Photo Preference task [28]. Based on these previous publications, we were able to select trials in each task that reflected high decision-difficulty or low decision-difficulty choices (see Figure 1B). This established a clear benchmark for replication: a particular device was sensitive to decision-difficulty if high decision-difficulty trials displayed significantly greater reaction time, movement time and trajectory curvature scores compared to low decision-difficulty trials [28, 13, 17].

In the Numeric-Size Congruity task, participants were asked to select which of two digits was larger in value, with the paired digits being either congruent in numeric and physical size (low decision-difficulty, e.g., 2 vs. 8) or incongruent in numeric and physical size (high decision-difficulty, e.g., 2 vs. 8). The Sentence Verification task asked participants to verify the truth of statements that could be non-negated (low decision-difficulty, e.g., ‘Cars have tires’) or negated (high decision-difficulty, e.g., ‘Cars do not have wings’). Finally, the Photo Preference task asked participants to select which of two dissimilarly-valenced (low decision-difficulty, e.g., High vs. Low pleasantness) or similarly-valenced (high decision-difficulty, e.g., High vs. High pleasantness) photos they preferred. Together, we ensured these tasks spanned a range of decision domains from objective perceptual judgments (e.g., digit discrimination), to semi-subjective conceptual judgements (e.g., truth value of a statement), and finally subjective preference judgements (e.g., preference for a particular photograph). These tasks also intentionally differed in stimulus characteristics (e.g., numeric, alphabetic, image), stimuli (e.g., numerical digits, written statements, photos), and processing requirements (e.g., perceptual discrimination, conceptual discrimination) allowing our results to be generalizable across remarkably distinct decision domains. Moreover, our experimental design allowed for a thorough exploration of the consistency of, and relationships between, metrics of decision-difficulty at different time points in the decision process (e.g., before and after movement-initiation). Finally, by building on previous mouse-tracking studies we are able to make strong a-priori predictions to provide a definitive test for using widely available touch-devices as a means of vastly extending the reach of decision science.

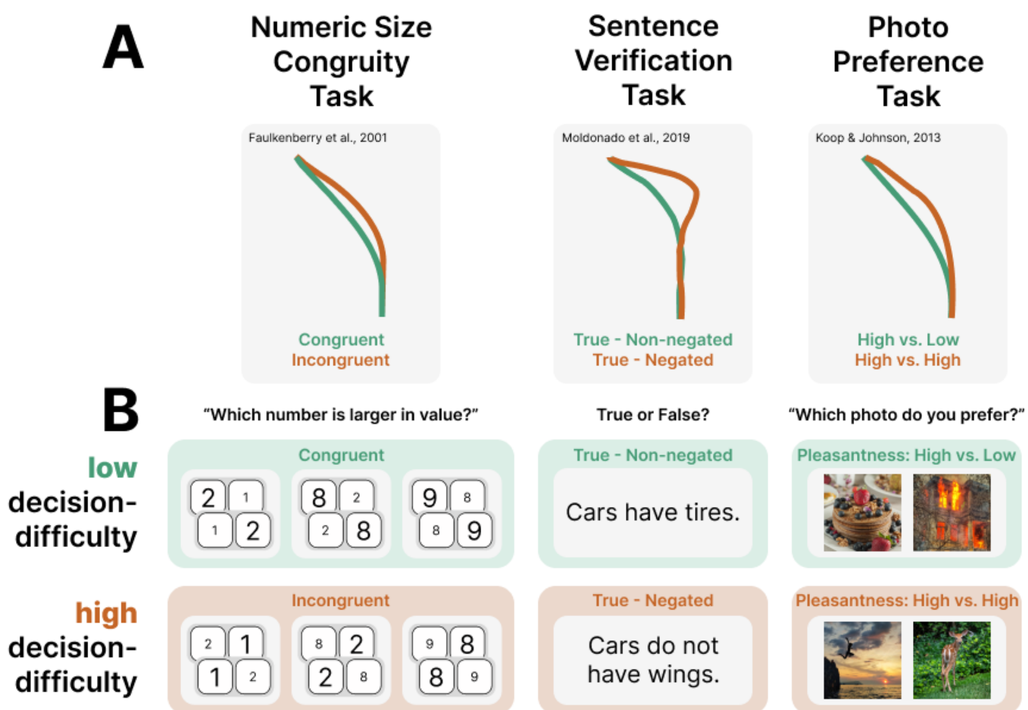


Figure 1: A) From left to right, a recreation of previous mouse trajectory results from the three task we replicate. Shown are average trajectories for the low (green) and high (orange) decision-difficulty categories for the Numeric-Size Congruity task (adapted from [17]), the Sentence Verification task (adapted from [31]'s replication of [13]), and the Photo Preference task (adapted from [28]). B) A representation of trial conditions falling within the low (green shading) and high (orange shading) decision-difficulty categories for each task, with stimuli examples.

2 Results

2.1 Tablets and smartphones measure decision-difficulty as well as computer mouse-tracking during reach-decision tasks

For all three tasks decision-difficulty was quantified as standardized reaction time, movement time and trajectory curvature (MAD) scores (see Methods subsection 5.3.2 - Dependent Measures). A replication of difficulty-driven effects was considered to have occurred should high decision-difficulty trials display significantly greater standardized scores than low decision-difficulty trials [28, 13, 17]. Thus, for each device (computer, tablet, smartphone) a-priori comparisons (t-tests) were made between high and low decision-difficulty trials within each task. A summary of statistics, unstandardized means, and mean differences between standardized scores are reported in Table 1.

For the Numeric-Size Congruity and Sentence Verification tasks, the paired samples t-tests replicated difficulty-driven for all three devices and for all three measures of decision-difficulty (see Table 1 and Figures 2 and 3). The Photo Preference task similarly replicated expected difficulty-driven effects across all measures during computer use, as well as for movement time and trajectory curvature during tablet and smartphone use (see Table 1 and Figure 4). Together, these results suggest that tablets and smartphones are sensitive tools for capturing information-rich reach-decision data across a variety of decision domains. Given the consistency of results for the other two tasks we attribute the divergence between computer and touch-device reaction time results during Photo Preference decisions to task features. Only the Photo Preference task required the judgment of a picture and we believe the fidelity of the picture information is degraded as screen-size is reduced, driving down the sensitivity to difficulty-driven effects on smaller displays. The relative increase in sensitivity to decision-difficulty for Computer reaction times is consistent with the Device differences described in the next Results subsection.

2.2 Mouse-tracking is more sensitive to decision-difficulty before movement while touch-device interactions are more sensitive during movement

Having established that all three devices tested capture decision-difficulty, our second analyses tested *how* the measurement of decision-difficulty changed across devices. Mean standardized reaction times, movement times and trajectory curvature scores for each task were separately submitted to a mixed-model ANOVA where we focused on main effects or interactions involving the between-subjects factor of Device factor and explored any (simple) main effects with pairwise comparisons between levels of Device (for results from this analysis outside this specific scope, including those that fully support the a-priori decision-difficulty effects described above, see Supplementary Materials 1). These tests revealed that the sensitivity of the specific metrics of decision-difficulty differed between touch-device and computer interactions. Specifically, computers showed increased sensitivity to decision-difficulty pre-movement (i.e., reaction time) while tablets and smartphones showed increased sensitivity during movement (i.e., movement time and trajectory curvature).

2.2.1 Measure sensitivity pre-movement

Within the Numeric-Size Congruity task, a 2 (Congruity) x 3 (Number Pairs) x 2 (Number Presentation Side) x 3 (Device) mixed-model ANOVA assessing standardized reaction times revealed both a main effect of Device ($F(2{,}237) = 12.69$, $p = 5.81\text{e-}6$, $\eta^2 = 3.16\text{e-}4$) and an interaction between Number Pair and Device ($F(4{,}237) = 14.23$, $p = 3.37\text{e-}10$, $\eta^2 = .022$). A significant main effect of Device was seen for both 1v2 ($F(2{,}237) = 17.79$, $p = 6.31\text{e-}8$) and 8v9 Number Pairings ($F(2{,}237) = 19.77$, $p = 1.15\text{e-}8$). The 8v9 effect, which is the hardest number-pair to decide between because it has both the smallest numeric difference and the smallest relative difference (see Supplementary Discussion 2), is driven by Computer having the longest reaction times compared to the touch-devices ($Mean_{Computer-Smartphone} = 0.18$, $t = 5.74$, $p = 6.01\text{e-}7$, $d = 0.43$; $Mean_{Computer-Tablet} = 0.20$, $t = 6.78$, $p = 1.30\text{e-}9$, $d = 0.50$). Meanwhile, the 1v2 effect, which is much easier because of the larger relative difference and presence of small numbers, is driven by Computer having the shortest reaction times ($M_{Computer-Smartphone} = -0.13$, $t = 4.26$, $p = 8.77\text{e-}4$, $d = 0.32$ and $M_{Computer-Tablet} = -0.16$, $t = 5.26$, $p = 7.74\text{e-}6$, $d = 0.39$). Thus, for reaction time, Computers show greater differentiation between hard and easy trials.

	Unstandardized						Standardized					
	M		*SD*				z_{Hard}-z_{Easy}					
	Decision Difficulty		Within		Between							
Device	**Easy**	**Hard**	**Easy**	**Hard**	**Easy**	**Hard**	*M*	*SE*	df	*t*	*p*	Cohen's *d*
Numeric-Size Congruity												
Reaction Time (ms)												
Computer	530.19	564.46	187.73	189.06	245.15	264.11	0.24	0.025	82	9.77	***	1.07
Tablet	556.12	571.19	127.02	124.66	149.23	152.21	0.16	0.026	78	6.14	***	0.69
Smartphone	503.24	526.08	121.35	135.78	169.16	184.01	0.18	0.034	77	5.15	***	0.58
Movement Time (ms)												
Computer	413.22	422.91	124.87	135.69	116.09	116.70	0.09	0.027	82	3.45	**	0.39
Tablet	513.31	538.46	85.38	102.68	135.83	146.86	0.27	0.029	78	9.35	***	1.05
Smartphone	469.64	498.98	89.62	109.12	120.13	119.29	0.31	0.029	77	10.55	***	1.20
Maximum Absolute Deviation (px)												
Computer	8.01	19.64	37.10	53.57	18.85	23.86	0.24	0.029	82	8.24	***	0.90
Tablet	13.04	25.75	27.70	38.18	10.55	13.42	0.38	0.030	78	12.51	***	1.41
Smartphone	12.90	26.92	30.72	41.87	8.93	14.10	0.37	0.027	77	13.94	***	1.58
Sentence Verification												
Reaction Time (ms)												
Computer	961.78	1496.46	326.40	493.04	395.28	631.25	1.15	0.043	82	26.57	***	2.92
Tablet	1013.20	1403.15	312.79	489.69	344.36	596.18	0.81	0.056	78	14.43	***	1.62
Smartphone	1041.42	1448.34	340.22	508.72	414.00	802.53	0.82	0.061	77	13.33	***	1.51
Movement Time (ms)												
Computer	462.91	606.26	174.68	274.51	164.11	257.50	0.52	0.050	82	10.50	***	1.15
Tablet	686.74	1056.88	215.76	413.78	241.59	631.90	0.79	0.055	78	14.26	***	1.61
Smartphone	627.03	995.52	199.39	413.78	210.84	491.94	0.91	0.050	77	18.40	***	2.08
Maximum Absolute Deviation (px)												
Computer	16.09	30.09	37.90	56.15	31.13	43.45	0.25	0.048	82	5.07	***	0.56
Tablet	16.70	35.64	27.31	36.12	27.79	42.43	0.44	0.054	78	8.13	***	0.91
Samrtphone	7.06	28.12	32.39	41.64	32.02	41.78	0.48	0.059	77	8.13	***	0.92
Photo Preference												
Reaction Time (ms)												
Computer	1024.93	1195.25	377.06	529.52	431.12	607.31	0.30	0.046	82	6.49	***	0.71
Tablet	1012.67	1048.80	389.61	379.82	454.26	576.58	0.04	0.048	78	0.80	n.s.	0.09
Smartphone	930.96	983.36	319.14	349.04	594.18	627.82	0.08	0.046	77	1.72	n.s.	0.20
Movement Time (ms)												
Computer	569.72	648.63	219.20	268.96	291.89	501.46	0.16	0.040	82	3.90	***	0.43
Tablet	782.13	895.75	235.46	325.69	298.36	398.20	0.23	0.047	78	4.95	***	0.56
Smartphone	722.06	796.08	231.83	308.75	258.66	373.29	0.16	0.044	77	3.28	*	0.37
Maximum Absolute Deviation (px)												
Computer	15.43	22.41	34.72	47.05	33.65	38.87	0.15	0.045	82	3.43	**	0.38
Tablet	20.90	30.89	33.85	36.08	21.71	20.35	0.32	0.057	78	5.87	***	0.64
Smartphone	24.16	31.056	32.61	38..31	25.06	21.39	0.15	0.054	77	2.78	*	0.32

Table 1: Task-specific unstandardized and z-scored means, and a-priori comparison results. Note. *p < .05; **p < .005; ***p < .0005

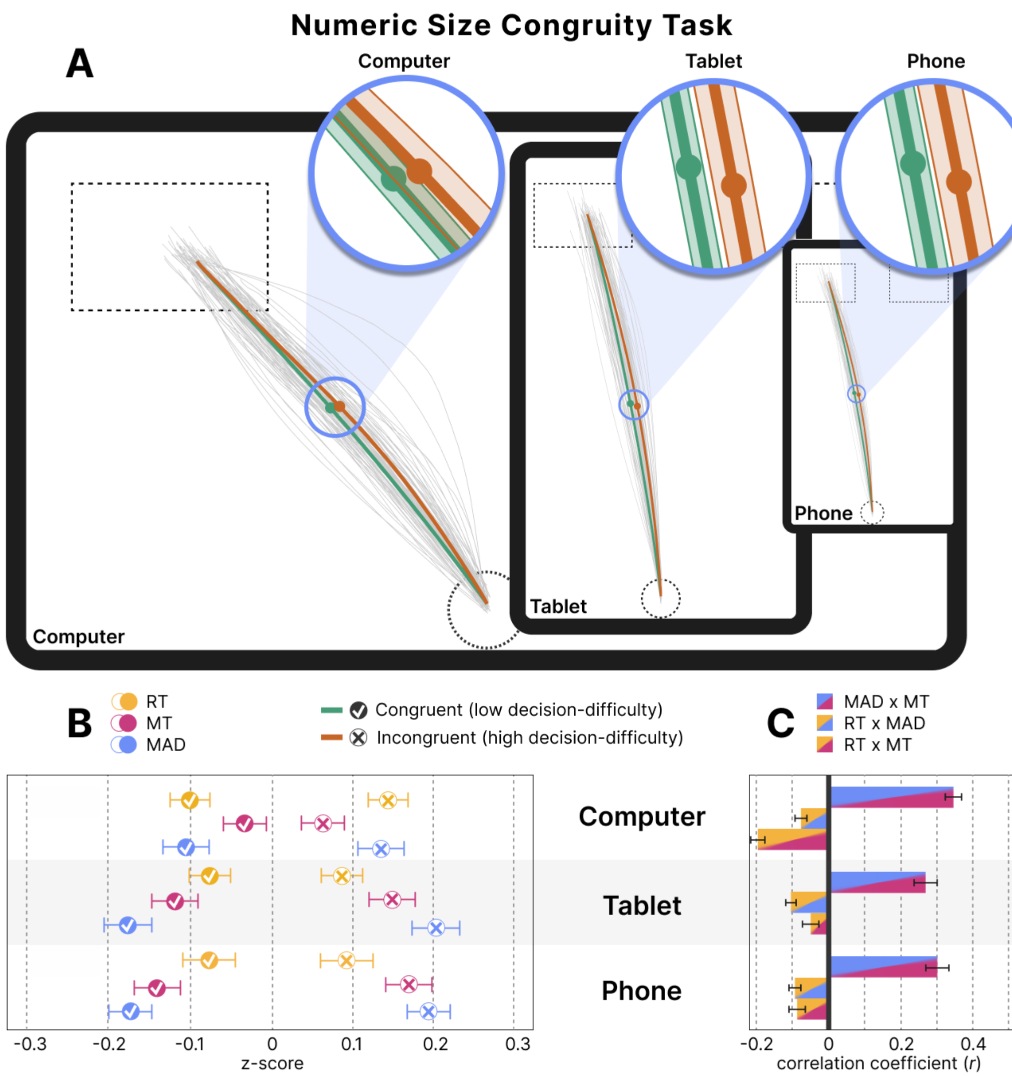


Figure 2: Numeric-Size Congruity task results. A) From left to right, trajectory results for computer, tablet and smartphone (phone) devices within screen size boundaries shown to scale of a representative physical device size. Light gray lines are each participants' average trajectory across all trials in this comparison. Mean trajectories across participants are shown for low (green line, Congruent trials) and high (orange line, Incongruent trials) decision-difficulty trials with the average location of maximum absolute deviation (MAD) shown with a filled circle. Insets zoom-in on the average point of MAD. Rightward reaches were mirrored to end left, and all reaches were space-normalized and standardized. Errors shown in the insets are the average of within-subjects standard error. For full trajectory visualization details, see Supplementary Note 1. B) From top to bottom, average of participant mean z-scored reaction times (yellow), movement times (pink), and maximum absolute deviation (blue) for computer, tablet and smartphone use. Error bars represent the averaged standard error of the difference between high and low difficulty means. C) Pearson's correlations (r) between measures of decision-difficulty for (from top to bottom) computer, tablet and smartphone use calculated from each participant and shown as an average. Error bars represent the standard error of the estimated marginal mean.

A similar pattern emerged in the Sentence Verification task. A 2 (Truth Value) x 2 (Polarity) x 3 (Device) mixed-model ANOVA revealed a three way interaction Truth x Negation x Device ($F(2{,}237)$ = 8.21, p = 3.57e-4, η^2 = .005) within reaction time. Based on where we predicted decision-difficulty to differ (see Figure 1) our follow-up tests looked at Negation x Device for True and False statements. We found a significant interaction only for True statements ($F(2{,}237)$ = 13.32, p = 3.30e-6, η^2 = .022). Breaking this down, Device was significant for both True-Negated statements ($F(2{,}238)$ = 8.22, p = 3.55e-4) and True-Non-negated statements ($F(2{,}238)$ = 14.27, p = 1.40e-6), but in importantly different ways. For the more difficult True-Negated statements, Computer reaction times were the longest ($M_{Computer-Tablet}$ = 0.16, t = 3.76, p = .003, d = 0.59; $M_{Computer-Smartphone}$ = 0.16, t = 3.82, p = .002, d = 0.60), but, for the easier True-Non-Negated statements, Computer reaction times were the shortest ($M_{Computer-Tablet}$ = -0.18, t = 4.25, p = 4.19e-4, d = 0.67; $M_{Computer-Smartphone}$ = -0.17, t = 4.11, p = 7.53e-4, d = 0.65). These results confirm that computers show greater differentiation across levels of decision-difficulty.

2.2.2 Measure sensitivity during-movement

An opposite pattern of results can be found when analyzing standardized movement time. Using the same ANOVA model described above, for Numeric-Size Congruity we found an interaction between Congruity and Device ($F(2{,}237)$ = 16.51, p = 1.93e-7, η^2 = .009). Follow-ups showed Device was significant for both Congruent ($F(2{,}237)$ = 18.15, p = 4.63e-8) and Incongruent trials ($F(2{,}237)$ = 14.22, p = 1.47e-6). Here, Computer showed *increased* movement times for Congruent trials ($M_{Computer-Smartphone}$ = 0.11, t = 5.38, p = 2.61e-6, d = 0.26; $M_{Computer-Tablet}$ = 0.088, t = 4.34, p = 3.06e-4, d = 0.21) but *decreased* movement times for Incongruent trials ($M_{Computer-Smartphone}$ = -0.11, t = 5.20, p = 6.08e-6, d = 0.21; $M_{Computer-Tablet}$ = -0.087, t = 4.30, p = 3.67e-4, d = 0.21), resulting in less divergence in movement times between the two difficulty levels compared to touch-devices. In complete opposition to the pattern observed for reaction times, these results suggest Computer movement times are significantly less sensitive to decision-difficulty compared to Tablet and Smartphone movement times.

Again Sentence Verification movement time results confirm this finding. Here the same task-specific mixed-model ANOVA described previously revealed a Negation by Device interaction ($F(2{,}237)$ = 19.59, p = 1.34e-8, η^2 = .027). Follow-ups revealed a main effect of Device both when statements were Non-negated ($F(2{,}237)$ = 21.43, p = 2.78e-9) and Negated ($F(2{,}237)$ = 16.82, p = 1.48e-7). Pairwise comparisons showed Computer having longer movement times compared to Tablets and Smartphones when statements were Non-negated ($M_{Computer-Smartphone}$ = 0.15, t = 5.76, p = 3.53e-7, d = 0.57; $M_{Computer-Tablet}$ = 0.12, t = 4.54, p = 1.33e-4, d = 0.44) and shorter movement times when statements were Negated ($M_{Computer-Smartphone}$ = -0.15, t = 5.96, p = 1.20e-7, d = 0.59; $M_{Computer-Tablet}$ = -0.11, t = 4.29, p = 3.85e-4, d = 0.42). This again results in less sensitivity in movement time between levels of Negation for the Computer condition compared to touch-devices.

The during-movement sensitivity observed for touch-devices also extended to trajectory curvature, but was impacted by the biomechanical properties of using a hand to act directly on a screen. Specifically, both tablet and smartphone results displayed a side of space biases where rightward reaches show more trajectory curvature compared to leftward reaches, matching what is observed in real reaching experiments [21]. Within Numeric-Size Congruity, this effect is evident in the trajectory curvature results as a Number Pair Presentation Side x Device interaction ($F(2{,}237)$ = 16.90, p = 1.38e-7, η^2 = .049) where both Left and Right reaches showed main effects of Device (Left: $F(2{,}237)$ = 17.07, p = 1.19e-7; Right: ($F(2{,}237)$ = 16.55, p = 1.86e-7), but in opposite directions. For Left reaches, Tablets and Smartphones show significantly less curvature than Computer trajectories ($M_{Computer-Tablet}$ = 0.27, t = 4.70, p = 6.47e-5, d = 0.52; $M_{Computer-Smartphone}$ = 0.30, t = 5.34, p = 3.30e-6, d = 0.59) while for Right reaches, Tablets and Smartphones show significantly more curvature than Computer trajectories ($M_{Computer-Tablet}$ = -0.26, t = 4.66, p = 7.96e-5, d = 0.51; $M_{Computer-Smartphone}$ = -0.29, t = 5.20, p = 6.57e-6, d = 0.57). Appreciating that Sentence Verification choice stimuli were locked to a side of space, the Sentence Verification trajectory curvature results bolster these directional effect findings, revealing a Truth x Device interaction ($F(2{,}237)$ = 15.16, p = 6.39e-7, η^2 = .074). Here we also see main effects of Device for both Left/True ($F(2{,}237)$ = 13.96, p = 1.86e-6) and Right/False reaches ($F(2{,}237)$ = 16.23, p = 22.47e-7) but in opposite directions. For Left/True reaches, Tablets and Smartphones show significantly less curvature than Computer trajectories ($M_{Computer-Tablet}$ = 0.25, t = 4.28, p = 4.06e-4, d = 0.59; $M_{Computer-Smartphone}$ = 0.30, t = 5.10, p = 1.03e-

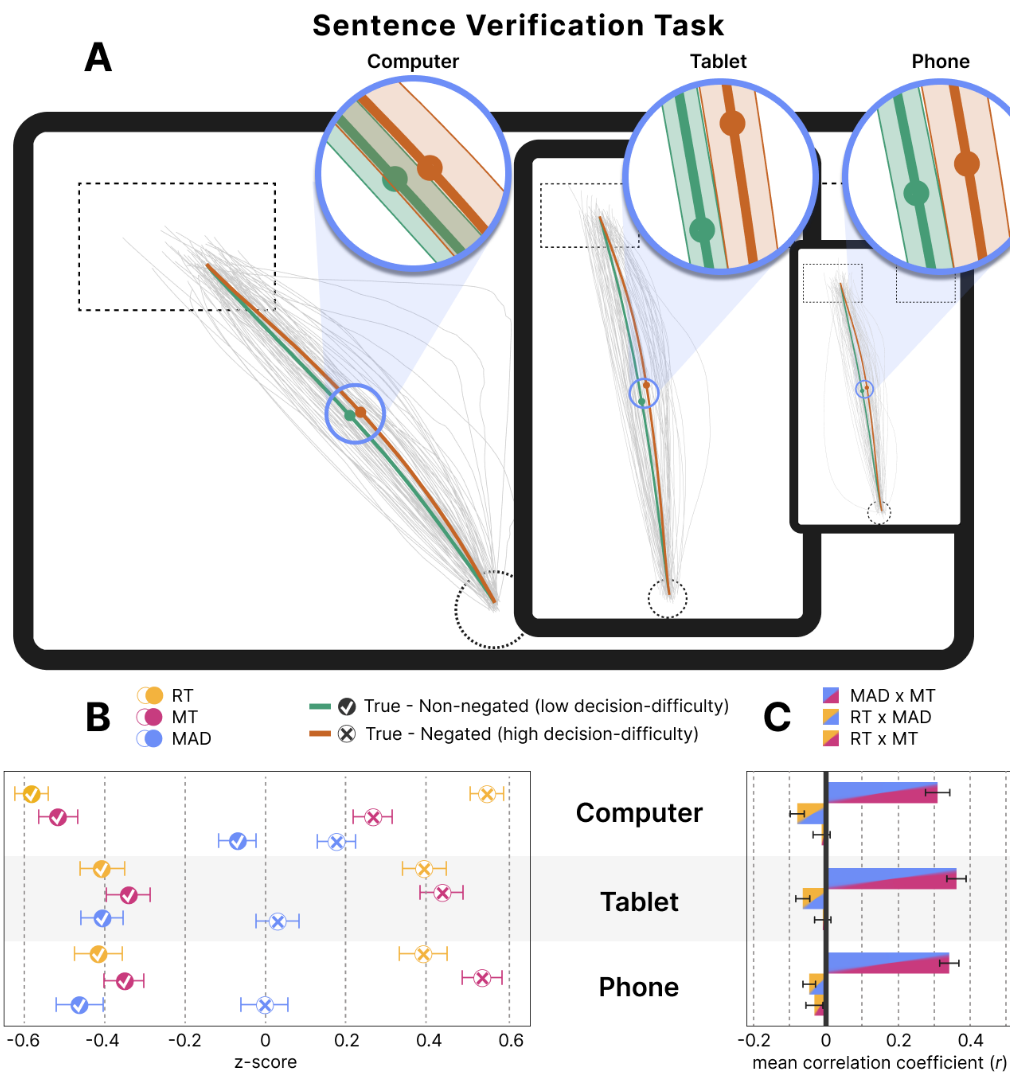


Figure 3: Sentence Verification task results. A) From left to right, trajectory results for computer, tablet and smartphone (phone) devices within screen size boundaries shown to scale of a representative physical device size. Light gray lines are each participants' average trajectory across all trials in this comparison. Mean trajectories across participants are shown for low (green line, True Non-negated trials) and high (orange line, True Negated trials) decision-difficulty trials with the average location of maximum absolute deviation (MAD) shown with a filled circle. Insets zoom-in on the average point of MAD. Rightward reaches were mirrored to end left, and all reaches were space-normalized and standardized. Errors shown in the insets are the average of within-subjects standard error. For full trajectory visualization details, see Supplementary Note 1. B) From top to bottom, average of participant mean z-scored reaction times (yellow), movement times (pink), and maximum absolute deviation (blue) for computer, tablet and smartphone use. Error bars represent the averaged standard error of the difference between high and low difficulty means. C) Pearson's correlations (r) between measures of decision-difficulty for (from top to bottom) computer, tablet and smartphone use calculated from each participant and shown as an average. Error bars represent the standard error of the estimated marginal mean.

5, d = 0.70) while for Right/False reaches, Tablets and Smartphones show significantly more curvature than Computer trajectories ($M_{Computer-Tablet}$ = -0.24, t = 4.18, p = 6.12e-4, d = 0.58; $M_{Computer-Smartphone}$ = -0.30, t = 5.09, p = 1.06e-5, d = 0.70). Again, this suggests that a right hand bias is more prominent for real touch interactions compared to mouse cursor movements (see Supplemental Discussion 3 for confirmatory evidence from the analysis of Movement Time).

Finally, the trajectory results from the Photo Preference task provide another example of how touch and mouse interactions differ. A 3 (Valence Pairing) x 3 (Device) mixed-model ANOVA revealed a main effect of Device (F(2,237) = 9.32, p = 1.27e-4, η^2 = .022) with standardized trajectory values for Computer responses (M = -0.0263, SD = 0.267) found to be different than Tablet (M = -0.116 , SD = 0.322; t = 3.50, p = .001, d = 0.31) and Smartphone responses (M = -0.132, SD = 0.036; t = 3.91, p = 3.64e-4, d = 0.35), and no significant difference between the two touch-devices. This Device effect did not significantly interact with decision-difficulty, indicating that this is a difference in the shape of the produced trajectories based on input - an idea which aligns with our interpretation that reaches produced as a result of direct interaction are different than those mediated by a mouse (see section 3 - Discussion). Overall, the differences in trajectory shape and presence of a right-hand bias in the Tablet and Smartphone results in contrast to Computer results point to a similarity between touch-device responses and real-world reaching when making choice selections. Further, these results highlight the increased sensitivity of post-movement measures during touch-device use.

2.3 Pre- and post-movement measures are flexible, non-redundant carriers of decision information

Here, we assess the relationship between our decision-difficulty measures to demonstrate that pre- and during-movement measures carry unique decision information. To do so, we obtained a within-participant correlation coefficient (r) for each combination of measures (Correlation-Type: $r_{MAD,MT}$ vs. $r_{MAD,RT}$ vs. $r_{MT,RT}$) within each task and device. These participant average correlation coefficients were then compared using a (3) Correlation Type x (3) Task x (3) Device mixed-model ANOVA. Where correlations between measures are positive, it would indicate that they carry redundant information. However, any inverse relationship would demonstrate a push and pull between measures showing that on any given trial, a best estimate of decision-difficulty should include both pre- and during-movement measures. The results of the ANOVA revealed a main effect of Task (F(2,237) = 22.06, p = 1.13e-9, η^2 = .009), a very strong main effect of Correlation-Type (F(2,237) = 601.10, p = 1.10e-92, η^2 = .45) and an interaction between Correlation-Type and Task (F(4,237) = 5.54, p = 6.47e-7, η^2 = .004). To follow up, we examined each Task separately and found a strong Correlation-Type effect in all three (SC: F(2,239) = 302.94, p = 2.85e-69, η^2 = .56; SV: F(2,239) = 242.55, p = 6.05e-53, η^2 = 0.50; PP: F(2,239) = 358.29, p = 6.13e-76, η^2 = .60). Mean r values revealed trajectory curvature and movement time ($r_{MAD,MT}$) to be moderately positively correlated (SC: M_r = 0.30, SD = 0.24; SV: M_r = 0.33, SD = 0.26; PP: M_r = 0.36, SD = 0.23) which intuitively makes sense - traveling a longer distance (MAD) usually takes a longer time (MT). In contrast, in each task, reaction time was found to be weakly inversely correlated with both other measures (SC: M_r = -0.092, SD = 0.14 and M_r = -0.11, SD = 0.20 for $r_{MAD,RT}$ and $r_{MT,RT}$ correlations, respectively; SV: M_r = -0.065, SD = 0.17 and M_r = 0.006, SD = 0.20 for $r_{MAD,RT}$ and $r_{MT,RT}$ correlations, respectively; PP: M_r = -0.065, SD = 0.15 and M_r = -0.041, SD = 0.19 for $r_{MAD,RT}$ and $r_{MT,RT}$ correlations, respectively). This pattern meant that the Correlation-Type comparisons always showed differences between during-movement correlations ($r_{MAD,MT}$, stronger and positive) and the pre- to during-movement correlations ($r_{MAD,RT}$ and $r_{MT,RT}$, weaker and negative). By task, the results of these pairwise comparisons were, for $r_{MAD,MT}$ vs. $r_{MAD,RT}$: SC: p = 3.5e-68, d = 2.00; SV: p = 1.06e-67, d = 1.86; PP: p = 8.23e-83, d = 2.19, and for $r_{MAD,MT}$ vs. $r_{MT,RT}$: SC: p = 2.18e-73, d = 2.11; SV: p = 2.15e-50, d = 1.53; PP: p = 2.04e-76, d = 2.07. The only slight difference across tasks we observed was that $r_{MT,RT}$ in the Sentence Verification task was close to zero, rather than weakly negative, and as such, there was a pairwise difference between $r_{MT,RT}$ and $r_{MAD,RT}$ (p = 7.70e-04, d = -0.33).

Taken together, this analysis reveals that pre- and during-movement measures display an intricate relationship independent of their role in indexing task-specific decision-difficulty. That is, while across all tasks and devices, reaction time, movement time and curvature increase with decision-difficulty (see Results subsection 2.1) on a trial-by-trial basis these measures adapt to the demands of the task and pre- and during-

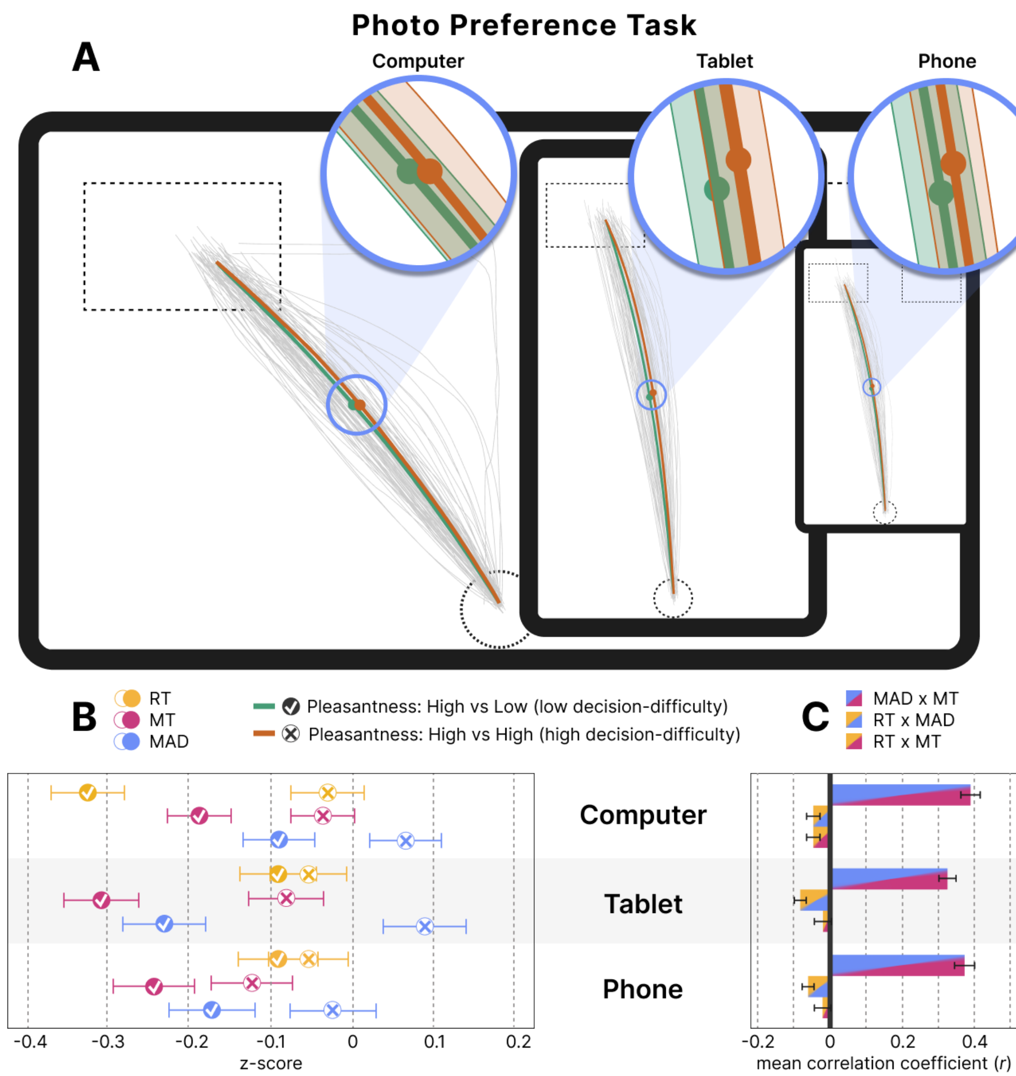


Figure 4: Photo Preference task results. A) From left to right, trajectory results for computer, tablet and smartphone (phone) devices within screen size boundaries shown to scale of a representative physical device size. Light gray lines are each participants' average trajectory across all trials in this comparison. Mean trajectories across participants are shown for low (green line, High vs. Low pleasantness trials) and high (orange line, High vs. High pleasantness trials) decision-difficulty trials with the average location of maximum absolute deviation (MAD) shown with a filled circle. Insets zoom-in on the average point of MAD. Rightward reaches were mirrored to end left, and all reaches were space-normalized and standardized. Errors shown in the insets are the average of within-subjects standard error. For full trajectory visualization details, see Supplementary Note 1. B) From top to bottom, average of participant mean z-scored reaction times (yellow), movement times (pink), and maximum absolute deviation (blue) for computer, tablet and smartphone use. Error bars represent the averaged standard error of the difference between high and low difficulty means. C) Pearson's correlations (r) between measures of decision-difficulty for (from top to bottom) computer, tablet and smartphone use calculated from each participant and shown as an average. Error bars represent the standard error of the estimated marginal mean.

movement measures function as non-redundant carriers of decision information. Specifically, it appears that on trials where participants react more quickly (shorter RTs) there is a slight increase in movement time and curvature (see section 3 - Discussion for further interpretation). It is also notable that there were no significant Device differences and limited differences due to Task. This highlights the remarkable stability both of this interplay between measures and for reach-decisions to track decision-difficulty across a variety of interface types.

3 Discussion

We investigated whether measuring reach decision-difficulty could be extended beyond computer use to tablets and smartphones through the deployment of a three-task online experiment across the three devices. Each task replicated a prior mouse-tracking study used to observe decision processes (Numeric-Size Congruity task [17], Sentence Verification task [13, 31], Photo Preference task [28]), allowing us to make strong predictions about which trials in each task would have high versus low decision-difficulty (see Figure 1).

Task-specific results replicated previous mouse-tracked outcomes, with high difficulty decisions displaying greater reaction times, movement times and trajectory curvature compared to low difficulty decisions. Most excitingly, all of these effects were replicated across all devices. Thus, this study demonstrates the robustness of dynamic measures of decision-making and offers validation for the use of small, portable devices to collect this movement information. For the Numeric-Size Congruity task [17], replication manifested as increased reaction time, movement time and trajectory curvature for incongruent trials compared to congruent trials (see Figure 2). For the Sentence Verification task [13, 31], the same metrics were increased on true-negated statements compared to true-non-negated statements (see Figure 3). Finally, for the Photo Preference task [28], movement time and trajectory curvature were increased for decisions requiring judgements between photos similar in pleasantness compared to decisions requiring judgements between photos dissimilar in pleasantness.

However, these a-priori comparisons also suggested that not all tasks might be suitable for deployment on smaller devices. Results from the Photo Preference task show that tablets and smartphones have a reduced sensitivity to decision-difficulty effects, especially for reaction time (see Table 1). We believe that this is a reflection of stimuli salience as screen size is reduced. While the other two tasks presented decision information as text, the Photo Preference task required participants to distinguish between two detailed photos, which likely degraded in stimulus information as the stimulus size decreased. Therefore, our key message is that all devices are able to track decision-difficulty but device differences exist and are important to understand. Our second cluster of results then specifically interrogated device differences. The results were clear: computer responses were consistently different from tablet and smartphone responses. Computer responses showed an increased sensitivity to decision-difficulty within pre-movement measures (reaction time) while touch-device responses revealed greater sensitivity during movement (movement time and trajectory curvature). We speculate this might be due to the different user-interaction requirements of touch-devices that enforce different 'reach' biomechanics compared to computer-mouse interactions. This is supported by the right-hand bias effects observed when swiping a finger/thumb or sliding a stylus but not when moving a mouse. This right-hand bias, also evident in real reaching [7, 21], is thought to arise from preferential processing of stimuli presented on the right of a display you are interacting with, resulting in less trajectory curvature and faster movement times during rightward reaches.

Why might smartphones and tablets show effects similar to a real reach movement? First, real-world movements made to enact mouse cursor changes on a screen are physically very small. While the cursor traverses a large on-screen distance, the hand moving the mouse travels a smaller distance in less time than even a finger on a smartphone (see non-standardized means in Table 1). These movements across less space and time produce more ballistic responses [24, 23]. As time and space during movement are at a premium with little of either available to express in indecision, this requires more of a decision to be resolved prior to movement initiation. [25, 52, 50]. The repercussions of front-loading the decision due to physical movement constraints align with results demonstrating that the demands of a motor task can directly influence cognitive processing (e.g., cognitive tuning [46, 10, 6, 33]). Here, it means that decision-differences arising in a computer task need to be more resolved prior to movement, leading to more sensitivity to difficulty being expressed by reaction time. More broadly, these results support the idea that the brain is optimized to take advantage of the affordances of the world it navigates, when more time and space are available because a physical movement is longer, the final commitment to a particular choice can be withheld well into movement execution [50].

A second explanation for the difference between pre- and during-movement sensitivity across computers compared to tablets and smartphones is the directness of the interaction. When moving a mouse to control a cursor to select a choice-option the action is physically dissociated from the target we are choosing - the hand is on the table rather than the screen. But, when we move our finger to touch a choice-option on a tablet or smartphone our action is directed toward the actual thing we are selecting. From the perspective of a brain controlling movement this is likely a profoundly different problem. For example, physically interacting with an object increases its appeal [51] and moving an object toward your own body can improve your ability to remember it [48]. These phenomena are likely related to the coordinate remapping required when moving a mouse in one plane to control a cursor in a different plane. This dramatically differs from the more direct planning available to the brain when mapping a touch screen target into the action space of the hand and arm [12, 53, 43, 49]. We would argue that it is this directness of interaction and movements that traverse longer distances over more time that explain why touch-devices show increased sensitivity in measures recorded during movement.

This dynamic interplay between pre- and during-movement measures was the subject of our third category of results. Despite all three measures increasing as decision-difficulty increased, our correlational analyses revealed an inverse relationship between reaction time and during-movement measures (also seen in some previous real-reaching tasks [16]). This discrepancy between overall task-related effects and trial-by-trial effects on the measures is compatible with an evidence accumulation framework of decision-making. Within this framework, evidence is noisily accumulated over time until a decision threshold has been reached [45, 50], signalling the onset of a movement. More difficult decisions require more evidence to be accumulated before support for one option reaches this threshold. This takes more time (i.e., longer reaction time), and unresolved competition impacts movements during choice selection (i.e., longer and less straight movements [47, 45, 50]), explaining the overall effects of decision-difficulty we report. However, when decision-difficulty is constant, there is still natural variation in reaction times. If decision processing requirements remain the same, but reaction time is reduced, there is more unresolved competition at movement onset. This necessarily shifts decision processes into the movement. As a result, on a trial-by-trial basis shorter reaction times will map to longer movement times and trajectories with more curvature - exactly the inverse relationship we report. Evidence accumulation thus accounts for both the a-priori main effects of decision-difficulty we report and the measure correlations we observe. Harder decisions result in increased reaction times, movement times and trajectory curvature because evidence accumulates more slowly in these cases. For any given decision where a set amount of evidence is required, however, there is a trade-off between pre-movement and during-movement decision resolution - abbreviating one elongates the other.

4 Conclusion

Across computers, tablets and smartphones, measured by reaction time, movement time and trajectory curvature, and capturing how these measures are dynamically related, reach-decision tasks provide a detailed read-out of decision-making. Given the ubiquitous use of touch-devices and websites, our validation of these metrics - across three diverse tasks and in a remote cohort of 240 participants - prove they are accessible outside the lab and impartial to the device used. The remarkable consistency of our results offers exciting new ways to apply these findings to research and industry, providing detailed knowledge of decision dynamics to domains such as corporate talent assessment and implicit bias measurement. Our results also offer the potential to optimize the collection of decision information, indicating that there are features of a decision and a device that make a certain combination the most sensitive for a particular task. Decisions and the movements we make to enact them literally shape our daily lives. By vastly expanding the accessibility of decision measures to include anyone with a touch-device we therefore hope to open new doors to the insights derived from this rich information.

To build on the incredible opportunity of remote data collection used to investigate the detailed dynamics of decision processes, we also conducted a second companion study (Bertrand et al., 2023). In this subsequent study we replicate the current study but integrate webcam eye-tracking, a technique which is sensitive to pre-movement decision processes. Together, this two-study series allows us a detailed description of the entirety of a decision - from the gaze deployed to gather information upon stimuli onset through the mouse-tracked movements produced to enact a final choice.

5 Methods

5.1 Participants

All experimental procedures were approved by the University of Alberta Research Ethics Office. 305 naive Amazon Mechanical Turk (www.mturk.com) participants took part in the study using either a computer, tablet or smartphone for a payment of $7 USD. Participation was restricted on Mechanical Turk to Canada- or U.S-based participants between 18 and 35 years of age who had an approval rating above 95% on 100 or more study completions. Participants self-reported age, gender, handedness, visual acuity, English language proficiency, habitual activities requiring hand-eye coordination, chosen device specifications and typical use of their chosen device for participation (see Supplementary Tables 1-3 for a complete demographic and device use summary). Participants were excluded from analysis based on insufficient ($< 50\%$) good trials within any of the experimental tasks or in any of the unique task conditions (see sub-subsection 5.3.3 - Data Cleaning).

5.1.1 Computer

101 participants completed the study using a personal computer. Of those, nine were excluded from analysis for not meeting device interaction requirements (i.e., did not use a wired or wireless mouse). A further nine computer users were excluded (see sub-subsection 5.3.3 - Data Cleaning), resulting in data from 83 computer users being analyzed (25 female, 56 male, and 2 who preferred not to say; $M_{age} = 33.75$, $SD_{age} = 9.35$).

5.1.2 Tablet

101 participants completed the study using a tablet. Four were excluded from analysis for not meeting device interaction requirements (i.e., did not use finger-, thumb- or stylus-based interactions). A further nineteen tablet users were excluded (see sub-subsection 5.3.3 - Data Cleaning), leaving data from 79 tablet users to be analyzed (27 female, 51 male, and 1 nonbinary; $M_{age} = 33.41$, $SD_{age} = 6.25$).

5.1.3 Smartphone

103 participants completed the study using a smartphone. Of those, twenty-five were excluded (see sub-subsection 5.3.3 - Data Cleaning), leaving 78 smartphone users for analysis (26 female, 52 male, and 1 who preferred not to say; $M_{age} = 33.73$, $SD_{age} = 6.72$).

5.2 Procedure and apparatus

The study was implemented using Labvanced [18], a graphical task builder offering built-in mouse- and finger-tracking, and temporal response recording compatible with computer, tablet and smartphone use for online study implementation. The study was distributed via Amazon Mechanical Turk, and devices used for study completion were uncontrolled except for requiring use of a separate mouse (wired or wireless) during computer use, or an Android operating system and touch-screen device interaction (via finger, thumb or stylus) during tablet or smartphone use (see Supplementary Tables 2-3 for selected device and interaction details).

Participants completed three reach-decision tasks requiring them to choose one of two stimuli presented at the top left and top right corners of their device screen based on a question or statement appearing at the center of the testing interface (see Figure 5). The reach-decision tasks (see Figure 1) presented Numeric-Size Congruity (adapted from [17]), Sentence Verification (adapted from [13, 31]) and Photo Preference (adapted from [28]) paradigms, each consisting of 84 trials and taking approximately 15 minutes to complete.

Each trial first presented a green circular start button labeled “Touch here” at the bottom center of the screen, requiring participants to navigate their mouse cursor to (Computer) or place their finger, thumb, or stylus on (Tablet and Smartphone) the button to start the trial. Touching the start button triggered a three second countdown, centered on the display screen (Figure 5B). Removing the mouse cursor, digit or stylus from the start button or the surface of the screen paused the countdown until touch-contact within the start button had been re-established. For the Numeric-Size Congruity and Photo Preference tasks, countdown onset was accompanied by a task-specific question appearing centered at the top of the display (Figure 5B). Upon countdown completion, two choice boxes appeared at the upper-left and upper-right of the screen, each presenting trial-specific choice options. For the Sentence Verification task, the two choice options appeared coincident with countdown onset and presented a statement centered at the top of the screen upon countdown completion (Figure 5B). Participants were free to select either choice option immediately upon countdown completion. For Computers, choice selection required participants to move their mouse cursor inside the choice-box. For Tablets and Smartphones, participants were required to slide their finger, thumb, or stylus across the screen to touch their selected choice-box, keeping contact with the screen at all times. If touchscreen contact was lifted, that trial was removed from analysis and an error message would appear on the screen, reading "Your finger was lifted from the screen as you moved, and we were unable to track the movement. Please touch your option now and remember in the future to keep your finger on the screen." When selected, a choice-box was highlighted with a blue border, the other option and start button disappeared, and a "Next" button appeared centered on the screen. Participants were then free to click or press on the "Next" button to continue to the next trial, allowing them to self-pace the experiment.

Trials were randomized within each task and the order of task presentation was counterbalanced across participants. Participants were instructed to complete the study in its entirety in a single session and were provided with detailed instructions outlining each task before it started. Participants were encouraged to take short breaks between tasks but had a maximum time limit of ninety minutes to complete the study.

Labvanced automatically scales the dimensions of the testing interface and its stimuli components to the screen size and resolution of the device in use, presenting a landscape (800 x 450 pixel, Labvanced coordinates) orientation for computer-based participation and a portrait (470 x 800 pixel, Labvanced coordinates) orientation for touch-device based participation. Stimuli-screen proportions remained consistent independent of device screen size (see Figure 5B for device-specific design details).

5.2.1 Numeric-Size Congruity

The Numeric-Size Congruity task in the current study was adapted from Faulkenberry, Cruise, Lavro and Shaki's experiment [17] examining the dynamics of the size congruity effect. For each Numeric-Size Congruity trial, the question "Which number is larger in value?" appeared coincident with the onset of the countdown timer, centered at the top of the screen (Figure 5). Following countdown termination two numbers were displayed simultaneously, one in each of the upper-left and upper-right choice boxes, and participants could move to select their preferred choice. Stimuli consisted of the Arabic numerals 1, 2, 8 and 9 displayed in Arial font and presented in pairs of different physical size with a 2:1 font size ratio. From these, six choice-pairs were generated: 1v2, 2v8 and 8v9, with each pair either congruent in physical and numeric size (the numerically larger numeral appearing physically larger than its paired counterpart, e.g., 2v8), or incongruent in physical and numeric size (the numerically larger numeral appearing physically smaller than its paired counterpart, e.g., 2v8; see Figure 1). Within each condition, the numerically larger number was presented equally often on the left and the right, counterbalancing side of space effects. This created twelve conditions, each presented 7 times for a total of 84 trials.

5.2.2 Sentence Verification

Adapted from Maldonado, Dunbar and Chemla's replication [31] of Dale and Duran's linguistic negation experiment [13], each Sentence Verification trial presented a "True" and "False" response option in the top-left and top-right corners of the screen, respectively (Figure 5). Following countdown termination, a statement was displayed at the top-center of the screen, prompting participants to judge whether it was true or false by selecting the appropriate response option. Statement stimuli consisted of 21 simple declarative statements manipulated in truth value (true, false) and negation (non-negated, negated). Sentence negation was manipulated by adding "not" to statements (e.g., "giraffes are tall" is non-negated, while "giraffes are not tall" is negated). Truth value was manipulated by changing the adjective at the end of the sentence (e.g., "giraffes are not short" is true, while "giraffes are not tall" is false). Crossing these factors yielded four sentence conditions where each sentence could be a true or false statement in either negated or non-negated

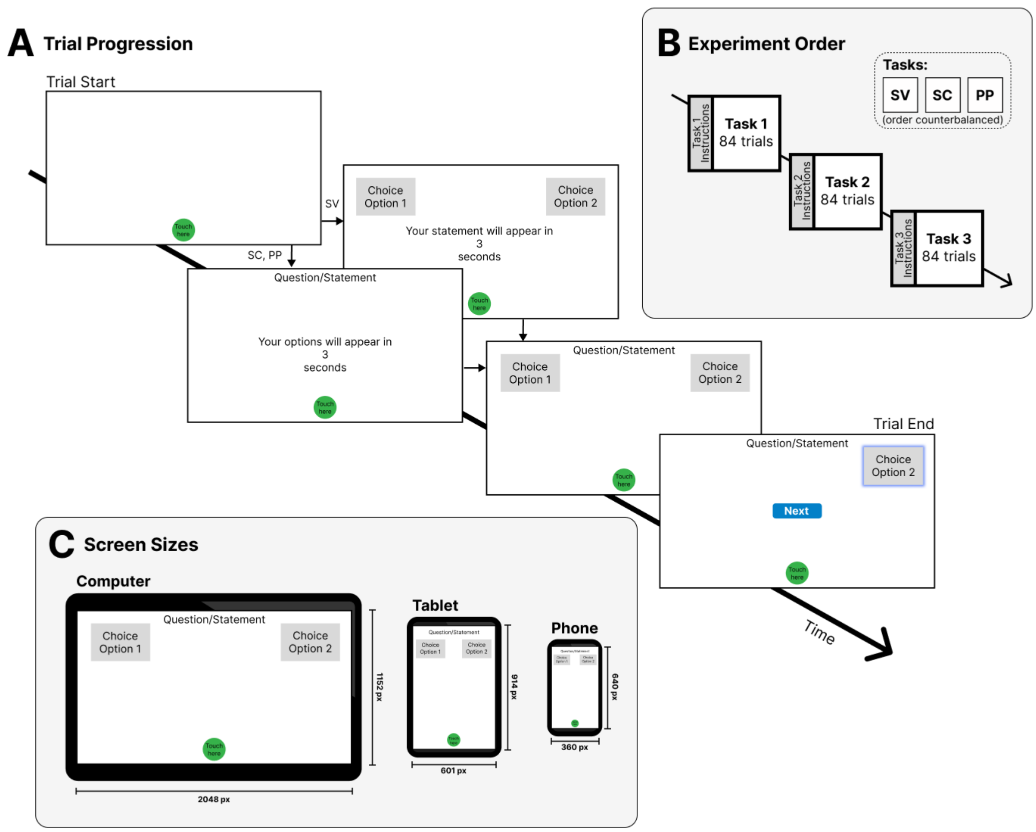


Figure 5: A) Overview of study design. Each participant completed a Numeric-Size Congruity task (SC), a Sentence Verification task (SV) and a Photo Preference Task (PP), with task order counterbalanced between participants. Task-specific instructions were presented prior to each task. B) All three tasks presented a classic reach-decision paradigm requiring participants to choose one of two stimuli presented at the top left and top right of their device screen. For SC and PP tasks, countdown onset was accompanied by a question specific to the task type appearing centered at the top of the display. The SV task presented the two choice options coincident with countdown onset and presented a statement (rather than a question) upon countdown completion. C) A comparison of interface arrangements between devices. Shown are representative examples of a computer, tablet and smartphone (phone) testing interface. All values are reported in pixels. Specific sizes of device screens and interface components observed by participants were dependent on the size of the device used, but screen to interface component proportions remained constant within each device category.

forms (see Figure 1 and Supplementary Table 4). Participants saw all four conditions of each statement, with the 84 resulting statements presented in a random order across trials.

5.2.3 Photo Preference

Adapted from Koop and Johnson's experiment [28] examining the mental dynamics of preferential choice, each Photo Preference trial presented the question "Which photo do you prefer?" centered at the top of the screen coincident with countdown initiation (Figure 5). Following countdown termination two images were then simultaneously displayed in the choice boxes to the upper left and upper right corners of the screen. As in Koop and Johnson [28], the International Affective Picture System (IAPS [29]) was used to develop a stimulus set of paired images using pleasantness ratings as an analog to photo preference, given equal levels of arousal [28]. We therefore selected 168 pictures from the IAPS, categorized as being high in pleasantness (pleasantness rating between 7 and 8), average in pleasantness (referred to as Med; pleasantness rating between 4.50 and 5.50) or low in pleasantness (pleasantness rating between 2 and 3). Images scoring greater than 6.15 in arousal were excluded. Selected pictures were then matched for arousal (difference < 0.30) and paired to create all pairwise combinations of High, Medium and Low. Pairs not matched in pleasantness (e.g., High−Med, High−Low, Med−Low) were counterbalanced for side of presentation, while pairs matched in pleasantness (e.g., High−High, Med−Med, Low−Low) appeared equally as often as the unmatched conditions when ignoring side of space (see Figure 1). This allowed for 14 presentations of each pleasantness pairing (7 of each unmatched pairing for each presentation side and 14 for matched pairings), for a total of 84 trials. Photo choice selections revealed a global preference for photos rated as more pleasant ($M_{MorePleasantSelected} = 78.3\%$), substantiating claims that preference is roughly analogous with pleasantness ratings [28]. As a result, the analysis included only trials containing a High pleasantness photo and in which the High photo was selected. Due to experimental error, half of participants completed a version of this task that did not counterbalance for side of presentation (i.e., High photos were always presented on the left). A separate ANOVA showed no significant difference between these groups for any measure, so both groups were included in the reported analysis where we collapsed across photo presentation side.

5.3 Data Treatment

5.3.1 Operationalization of trajectory data

Raw movement data was resampled to 60 Hz, then filtered using a 10 Hz lowpass filter. Reach onset was defined as the first time the mouse cursor (Computer) or finger/thumb/stylus (Tablet and Smartphone) ascended to 5% of its peak velocity within the start button and after countdown had terminated. Should this velocity threshold not be achieved prior to leaving the start button, this threshold was iteratively reduced by 5% until a reach onset could be defined. Reach offset was similarly defined as the first time the mouse cursor (Computer) or finger/thumb/stylus (Tablet and Smartphone) velocity descended below a velocity threshold of 5% peak velocity while within one of the two choice option boxes, with this threshold iteratively increasing by 5% if necessary.

5.3.2 Dependent Measures

For each trial, the following behavioural measures were obtained:
Reaction time (seconds): time from countdown termination to reach onset.
Movement time (seconds): time from reach onset to reach offset (choice selection).
Trajectory curvature (MAD): Within each trial, the perpendicular distance of the observed trajectory relative to a straight line connecting the trajectory start and end positions was calculated for each data point. Maximum absolute deviation (MAD) reports the maximum of these perpendicular distances. Straight trajectories produce values approaching zero while those curving toward the center of the screen were assigned positive MAD values and those moving away from the center were assigned negative MAD values.

Within-participant and within-task z-scores were computed for each dependent measure (reaction time, movement time, trajectory curvature). This standardization of within-participant measures allows for between-task and between-participant comparisons while controlling for participant variability and individual reach patterns. All analyses were conducted on these standardized values. See Table 1 for reporting of raw and standardized measure values.

5.3.3 Data cleaning

Data cleaning processes were identical independent of device and were conducted using customized MATLAB scripts. Errors on each trial could be a combination of reaches with recording errors, reaches with insufficient data points (fewer than seven unique positions), reaches with reaction times greater than 0.1s, reaches with movement times > 3 SD above a participants mean movement time, and reaches with reaction times > 3 SD above a participants mean reaction time. For Numeric-Size Congruity and Sentence Verification tasks, incorrect trials were also removed from analysis. As these tasks previously demonstrated very high levels of accuracy [13, 17], incorrect responses were considered to arise from participant error, with sustained performance errors indicating participant unreliability. The average percentage of total participant trials falling within each of these error categories are reported in Supplementary Table 5. A participant was excluded from analysis if, after data cleaning, they failed to have at least four trials in each condition of analysis as reported per task. In total, participants whose data was included for analysis had a mean of 95.6% usable trials for analysis (Range: 83.7%−98.4%).

5.4 Analysis

The main objective of this analysis was to determine whether task-specific decision-difficulty effects (as expected by previous studies, e.g., [17, 28, 13, 31]) were replicated and whether these effects were consistent despite differences in testing device. To that end, analysis proceeded in three primary stages: 1) a-priori comparisons to determine replication of antecedent results, 2) within-task, between-device omnibus analysis of variances (ANOVAs) to determine any effects or interactions arising due to device differences, and 3) between-device ANOVA to determine whether there are correlational relationships between measures of decision-difficulty and if these remain consistent across device.

5.4.1 A-Priori Comparison Procedure

To determine replication of the previous task-specific difficulty effects, a subset of trial conditions were selected to represent low and high difficulty decisions within each task (see Figure 1). For the Numeric-Size Congruity task, decision-difficulty followed size-congruity, with trials incongruent in numerical and physical size categorized as high in decision-difficulty, while congruent trials were categorized as low in decision-difficulty [17]. For the Sentence Verification task, decision-difficulty varied according to negation, with true statements the greatest negation-driven effects [13]. The current study therefore categorized true negated trials as representative of a high difficulty decision, and true non-negated trials as having low decision-difficulty. Finally, decision-difficulty in the Photo Preference task was driven by the similarity in pleasantness between photos [28]. The current study places trials comparing two photos high in pleasantness in the high decision-difficulty category, and trials comparing a photo high in pleasantness and one low in pleasantness in the low decision-difficulty category.

Within each task, mean standardized reaction time, movement time and trajectory curvature scores for low and high decision-difficulty trials were compared using a paired t-test. As these were a-priori tests based on replicating known effects, significance was set to $p \leq .05$ with no correction for multiple comparisons.

5.4.2 Within-task ANOVA Procedure

Mean standardized reaction time, movement time and maximum absolute deviation measures were separately submitted to mixed-model ANOVAs, with within-subject factors determined by individual tasks design and between-subject factors of device (computer, tablet, smartphone, see section 2 - Results). All multi-way mixed- and RM-ANOVAs were family-wise error corrected using a sequential Bonferroni procedure [11], and all repeated-measures main effects and interactions were Greenhouse-Geiser corrected to protect against violations of sphericity. The primary objective of this series of tests was to look for device differences. As a result, here we focus only on main effects or interactions involving Device. Full results outside this explicit objective can be found in Supplementary Materials 1, including results that support the a-priori tests of decision-difficulty. Interactions involving Device first collapsed over factors that did not interact, then were followed up by separating by the factor(s) other than Device. Significant (simple) main effects of Device were explored with all possible pairwise comparisons which were Bonferroni corrected with significance set at a corrected $p \leq .01$.

5.4.3 Between-task ANOVA Procedure

To explore the relationship between measures of decision-difficulty, a Pearson's correlation coefficient (r) was calculated between each pair of measures ($r_{MAD,MT}$, $r_{MAD,RT}$ and $r_{MT,RT}$) indicating the direction and strength of the relationship across trials for each participant within each condition, task, and device. Mean correlation coefficients were then submitted to a mixed-model ANOVA with Correlation-type and Task as within-subjects factors and Device as a between-subjects factor. Corrections and follow-up procedures were then conducted as described above, except here we were most interested in the pairwise comparisons between Task.

Acknowledgments

This study was funded by an NSERC Discovery Grant to CSC, and by an NSERC CGS-M and MITACS Accelerate International award to AOZ.

References

[1] Herman Aguinis, Isabel Villamor, and Ravi S Ramani. "MTurk research: Review and recommendations". In: *Journal of Management* 47.4 (2021), pp. 823–837.

[2] Alexander Anwyl-Irvine et al. "Realistic precision and accuracy of online experiment platforms, web browsers, and devices". In: *Behavior research methods* 53.4 (2021), pp. 1407–1425.

[3] Alexander L Anwyl-Irvine et al. "Gorilla in our midst: An online behavioral experiment builder". In: *Behavior research methods* 52.1 (2020), pp. 388–407.

[4] Pavlo Bazilinskyy and Joost C. F. de Winter. "Crowdsourced Measurement of Reaction Times to Audiovisual Stimuli With Various Degrees of Asynchrony". In: *Human Factors* (2018). DOI: 10.1177/0018720818787126.

[5] Daniel Bernoulli. "Exposition of a New Theory on the Measurement of Risk". In: *Econometrica* (1954). DOI: 10.2307/1909829.

[6] Diana Burk et al. "Motor effort alters changes of mind in sensorimotor decision making." In: *PLOS ONE* (2014). DOI: 10.1371/journal.pone.0092681.

[7] Craig S. Chapman et al. "Reaching for the unknown: multiple target encoding and real-time decision-making in a rapid reach task". In: *Cognition* (2010). DOI: 10.1016/j.cognition.2010.04.008.

[8] Craig S. Chapman et al. "Short-term motor plasticity revealed in a visuomotor decision-making task." In: *Behavioural Brain Research* (2010). DOI: 10.1016/j.bbr.2010.05.012.

[9] Paul Cisek and John F. Kalaska. "Neural Mechanisms for Interacting with a World Full of Action Choices". In: *Annual Review of Neuroscience* (2010). DOI: 10.1146/annurev.neuro.051508.135409.

[10] Ignasi Cos, Farid Medleg, and Paul Cisek. "The modulatory influence of end-point controllability on decisions between actions." In: *Journal of Neurophysiology* (2012). DOI: 10.1152/jn.00081.2012.

[11] Angélique O. J. Cramer et al. "Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies." In: *Psychonomic Bulletin & Review* (2016). DOI: 10.3758/s13423-015-0913-5.

[12] Helen A. Cunningham and Robert B. Welch. "Multiple concurrent visual-motor mappings: implications for models of adaptation." In: *Journal of Experimental Psychology: Human Perception and Performance* (1994). DOI: 10.1037/0096-1523.20.5.987.

[13] Rick Dale and Nicholas D. Duran. "The Cognitive Dynamics of Negated Sentence Verification". In: *Cognitive Science* (2011). DOI: 10.1111/j.1551-6709.2010.01164.x.

[14] Dror Dotan et al. "On-line confidence monitoring during decision making". In: *Cognition* (2018). DOI: 10.1016/j.cognition.2017.11.001.

[15] Dror Dotan et al. "Track It to Crack It: Dissecting Processing Stages with Finger Tracking." In: *Trends in Cognitive Sciences* (2019). DOI: 10.1016/j.tics.2019.10.002.

[16] Christopher D. Erb, Jeff Moher, and Stuart Marcovitch. "Attentional capture in goal-directed action during childhood, adolescence, and early adulthood." In: *Journal of Experimental Child Psychology* (2022). DOI: 10.1016/j.jecp.2021.105273.

[17] Thomas J. Faulkenberry et al. "Response trajectories capture the continuous dynamics of the size congruity effect". In: *Acta Psychologica* (2016). DOI: 10.1016/j.actpsy.2015.11.010.

[18] Holger Finger et al. "LabVanced: a unified JavaScript framework for online studies". In: *International conference on computational social science (cologne)*. 2017.

[19] Michael C. Frank et al. "Using Tablets to Collect Data From Young Children". In: *Journal of Cognition and Development* (2016). DOI: 10.1080/15248372.2015.1061528.

[20] Jonathan B. Freeman. "Doing Psychological Science by Hand". In: *Current Directions in Psychological Science* (2018). DOI: 10.1177/0963721417746793.

[21] Jason P. Gallivan and Craig S. Chapman. "Three-dimensional reach trajectories as a probe of real-time decision-making between multiple competing targets." In: *Frontiers in Neuroscience* (2014). DOI: 10.3389/fnins.2014.00215.

[22] Jason P. Gallivan et al. "Decision-making in sensorimotor control." In: *Nature Reviews Neuroscience* (2018). DOI: 10.1038/s41583-018-0045-9.

[23] Claude Ghez et al. "Discrete and continuous planning of hand movements and isometric force trajectories". In: *Experimental Brain Research* (1997). DOI: 10.1007/pl00005692.

[24] Claude Ghez et al. "Roles of proprioceptive input in the programming of arm trajectories." In: *Cold Spring Harbor Symposia on Quantitative Biology* (1990). DOI: 10.1101/sqb.1990.055.01.079.

[25] Adrian M. Haith, David M. Huberdeau, and John W. Krakauer. "Hedging your bets: intermediate movements as optimal behavior in the context of an incomplete decision." In: *PLOS Computational Biology* (2015). DOI: 10.1371/journal.pcbi.1004171.

[26] Eric Hehman, Ryan M. Stolier, and Jonathan B. Freeman. "Advanced mouse-tracking analytic techniques for enhancing psychological science". In: *Group Processes & Intergroup Relations* (2015). DOI: 10.1177/1368430214538325.

[27] Mel W. Khaw, Paul W. Glimcher, and Kenway Louie. "Normalized value coding explains dynamic adaptation in the human valuation process." In: *Proceedings of the National Academy of Sciences of the United States of America* (2017). DOI: 10.1073/pnas.1715293114.

[28] Gregory J. Koop and Joseph G. Johnson. "The response dynamics of preferential choice." In: *Cognitive Psychology* (2013). DOI: 10.1016/j.cogpsych.2013.09.001.

[29] Pj Lang. "International affective picture system (IAPS) : affective ratings of pictures and instruction manual". In: *CTIT technical reports series* (2005). DOI: null.

[30] Joshua de Leeuw. "JsPsych: a JavaScript library for creating behavioral experiments in a Web browser." In: *Behavior Research Methods* (2015). DOI: 10.3758/s13428-014-0458-y.

[31] Mora Maldonado, Ewan Dunbar, and Emmanuel Chemla. "Mouse tracking as a window into decision making." In: *Behavior Research Methods* (2019). DOI: 10.3758/s13428-018-01194-x.

[32] Gregory McCarthy and Emanuel Donchin. "A metric for thought: a comparison of P300 latency and reaction time". In: *Science* (1981). DOI: 10.1126/science.7444452.

[33] Jeff Moher and Joo-Hyun Song. "Perceptual decision processes flexibly adapt to avoid change-of-mind motor costs." In: *Journal of Vision* (2014). DOI: 10.1167/14.8.1.

[34] Camillo Padoa-Schioppa and John A. Assad. "Neurons in the orbitofrontal cortex encode economic value". In: *Nature* (2006). DOI: 10.1038/nature04676.

[35] Stefan Palan and Christian Schitter. "Prolific.ac—A subject pool for online experiments". In: *Journal of Behavioral and Experimental Finance* (2017). DOI: 10.1016/j.jbef.2017.12.004.

[36] John Palmer, Alexander C. Huk, and Michael N. Shadlen. "The effect of stimulus strength on the speed and accuracy of a perceptual decision". In: *Journal of Vision* (2005). DOI: 10.1167/5.5.1.

[37] Eliza Passell et al. "Cognitive test scores vary with choice of personal digital device". In: *Behavior Research Methods* (2021). DOI: 10.3758/s13428-021-01597-3.

[38] John W. Payne. "Task complexity and contingent processing in decision making: An information search and protocol analysis". In: *Organizational Behavior and Human Performance* (1976). DOI: 10.1016/0030-5073(76)90022-2.

[39] Thomas Pronk et al. "Can we measure individual differences in cognitive measures reliably via smartphones? A comparison of the flanker effect across device types and samples". In: *Behavior Research Methods* (2022), pp. 1–12.

[40] Antonio Rangel and Todd A. Hare. "Neural computations associated with goal-directed choice". In: *Current Opinion in Neurobiology* (2010). DOI: 10.1016/j.conb.2010.03.001.

[41] Michael Schulte-Mecklenbeck et al. "Process-tracing methods in decision making: on growing up in the 70s". In: *Current Directions in Psychological Science* (2017). DOI: 10.1177/0963721417708229.

[42] Kilian Semmelmann et al. "U Can Touch This: How Tablets Can Be Used to Study Cognitive Development". In: *Frontiers in Psychology* (2016). DOI: 10.3389/fpsyg.2016.01021.

[43] Britne A. Shabbott and Robert L. Sainburg. "Learning a visuomotor rotation: simultaneous visual and proprioceptive information is crucial for visuomotor remapping". In: *Experimental Brain Research* (2010). DOI: 10.1007/s00221-010-2209-3.

[44] Paul E. Stillman, Xi Shen, and Melissa J. Ferguson. "How Mouse-tracking Can Advance Social Cognitive Theory". In: *Trends in Cognitive Sciences* (2018). DOI: 10.1016/j.tics.2018.03.012.

[45] Paul E. Stillman et al. "Using dynamic monitoring of choices to predict and understand risk preferences." In: *Proceedings of the National Academy of Sciences of the United States of America* (2020). DOI: 10.1073/pnas.2010056117.

[46] Fritz Strack, Leonard L. Martin, and Sabine Stepper. "Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis." In: *Journal of Personality and Social Psychology* (1988). DOI: 10.1037/0022-3514.54.5.768.

[47] Nicolette Sullivan et al. "Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed". In: *Psychological science* 26.2 (2015), pp. 122–134.

[48] Grace Truong et al. "Mine in motion: how physical actions impact the psychological sense of object ownership". In: *Journal of Experimental Psychology: Human Perception and Performance* (2016). DOI: 10.1037/xhp0000142.

[49] Kunlin Wei et al. "Computer Use Changes Generalization of Movement Learning". In: *Current Biology* (2014). DOI: 10.1016/j.cub.2013.11.012.

[50] Nathan J. Wispinski, Jason P. Gallivan, and Craig S. Chapman. "Models, movements, and minds: bridging the gap between decision making and action". In: *Annals of the New York Academy of Sciences* (2020). DOI: 10.1111/nyas.13973.

[51] Nathan J. Wispinski et al. "Reaching reveals that best-versus-rest processing contributes to biased decision making." In: *Acta Psychologica* (2017). DOI: 10.1016/j.actpsy.2017.03.006.

[52] Aaron L. Wong and Adrian M. Haith. "Motor planning flexibly optimizes performance under uncertainty about task goals." In: *Nature Communications* (2017). DOI: 10.1038/ncomms14624.

[53] Kenji Yamamoto et al. "Rapid and long-lasting plasticity of input-output mapping." In: *Journal of Neurophysiology* (2006). DOI: 10.1152/jn.00209.2006.

Appendix B: Supplemental Materials from Chapter 2 Publication (Bertrand & Chapman, 2023)

B.1 README

README - Supplementary Materials

! **For the purposes of my thesis, the zip folder contents can be accessed in a Google Drive folder located here:**

https://drive.google.com/drive/folders/1WVIQ-PwAacwaMGgOI0LrfZumqhhYb7Sm?usp=sharing

(Please copy and paste link into browser - embedding not possible with combined PDFs in LaTeX)

In this zip folder, the following supplementary materials are included:

1.0 - Prolific Project Description
- This is a screenshot, in PDF form, of the description, including requirements and hardware information, that is provided to participants prior to accepting participation in our study. We provide this information in our supplemental materials as we believe, after extensive piloting, that this level of transparency and detail is important for minimizing participant and experimenter frustration or confusion.

2.0 - Instructional Video
- This mp4 file is shown to participants as part of the task instructions. Because we require participants complete an eight-movement sequence in a particular order, we believe including a video for participants to best understand the task was useful. This is shown to participants during a series of instructions and is not the only information they are provided with.

3.0 - Data Processing Diagram
- This PDF file is a detailed visual representation of the data processing procedure, including our approach to gaze-data-driven AOI clusters, trial and participant rejection criteria, and our segmentation approach. This diagram also includes the number of participants rejected at each step, showing how we went from 51 datasets collected to the 29 datasets included in our analysis.

4.0 - Follow-up Pairwise Comparisons - Supplementary Statistical Tests
- This PDF file is a text document that shares, in detail, the statistics associated with the pairwise results that were only summarized in the main text. We also share the alternative follow-up approach to the Fixation Duration measure (for information purposes only), to highlight the potential for nuanced spatial biases in our eye-tracking data.

5.0 - Spatial Assessment of Clustering Process
- This PDF file is a detailed visual representation of the process followed to assess the spatial qualities of the generated clusters, and also includes two figures that visualize the findings of this assessment. This spatial assessment is meant to complement the analogous temporal analysis shown in Figure 3 in the main text.

6.0 - Labvanced Study Link
- We share, in a PDF file, the link for the Labvanced study within the Labvanced library. The study is for demonstration purposes only, and is able to be imported into a free account to explore the back-end, experimenter view of our task, or, without creating an account, is available to view as a participant. Please note the warning at the outset that the participant view is for demonstration purposes only (i.e. no payment information applies, and no data will be used).

B.2 Prolific Project Description

1.0 - Prolific Project Description

Choose between 2 presented options -with webcam-based eye-tracking (*need high-perf. computer, graphics card+internet*)

By [redacted]

£4.50 • £6.00/hr 45 mins 3 places

Welcome to the experiment! This study should take approximately 45 minutes, and needs to be completed in one sitting. You will be required to stay **VERY STILL** during the entirety of the experiment for eye-tracking purposes. **Some people may not have the internet speed or graphics card + hard drive system performance that this study requires. If you've closed other internet tabs etc. and you still get an error that says the face/video processing is not working well enough, you will not get past the very first screen. You will be unable to participate and will need to return the study. Do not keep trying - it will not work. Thank you!**

We are conducting an academic study about decision-making. To study this, we need participants to choose between two presented options on the computer while a webcam-based eye-tracking system determines where the participants eyes are looking. While you complete this decision task, your eye gaze position will be recorded. While the webcam is used to calculate your eye gaze position in numerical form, no webcam video data is recorded.

Warning: some images shown may contain potentially explicit or offensive content (e.g. blood, violence) that some viewers may find disturbing. Participant discretion is advised.

We appreciate your interest in participating! This scientific data will be used to help us develop a better understanding of the eye-gaze in decision-making.

Still interested? Here's the important details:

Very important - please read the following requirements carefully

To participate, make sure you:

-- Use Chrome as your browser

-- Use a computer MOUSE (NOT A TRACKPAD OR LAPTOP TOUCHPAD - WILL FAIL)

-- Use only a laptop or desktop computer (tablets/phones will fail)

-- Can do a computer task without needing to wear glasses (contacts OK)

-- Have a connected webcam (either built-in or external device is fine)

-- Have your laptop or desktop computer on a stable table top (not on your lap), in a well-lit room - with light in front of you (not behind) so that your face isn't shadowed. Cannot be in a dark room with only light coming from the computer screen (eye-tracking system will fail)

-- Are aged 18 to 35

-- Have the ability to keep a consistent, still head position for the full duration of the study

-- Can keep the study full screen for the full study duration (approx. 45 mins)

-- Close all other internet tabs before starting the study (experiment will fail)

The task will take approximately 45 minutes to complete. Complete the task in one sitting. You will not receive payment if you do not complete the task fully and in one sitting.

DO NOT exit from full screen mode during the experiment. If you are not using the Chrome browser, please swap browsers before beginning the study. The study will take 2-3 minutes to load at the beginning of the study.

VERY IMPORTANT: Please close all tabs before beginning the study. You will receive an error if your background activity is impeding the performance requirements of the eye-tracking system and you'll need to return the study if the performance threshold is not met at the beginning of the study (you will get an error screen that tells you this). Some people may not have the internet speed or system performance that this study requires. If you've closed other internet tabs etc. and you still get an error that says the face/video processing is not working well enough, you will not get past the very first screen. You will be unable to participate and will need to return the study - do not keep trying. Thank you.

This research is carried out by researchers from [redacted] The plan for this research has been reviewed for its adherence to ethical guidelines and approved by Research Ethics Board [redacted]

Devices you can use to take this study:

Desktop

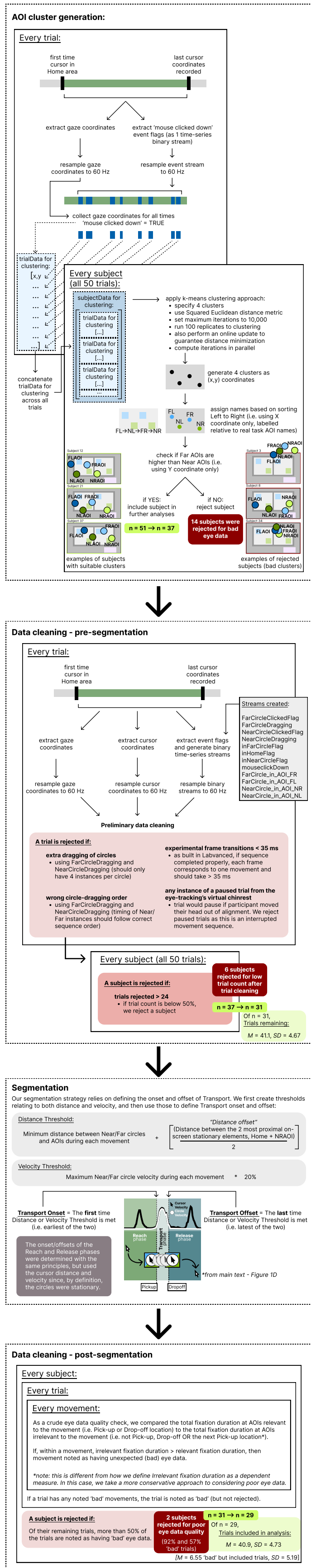
You will also need:

Camera

Open study link in a new window

B.3 Data Processing Pipeline

3.0 - Data Processing Diagram



B.4 Follow-up Pairwise Comparisons

4.0 - Follow-up Pairwise Comparisons - Supplementary Statistical Tests

Fixation Duration

Our primary comparison of interest, as described in the main text, was between Relevant and Irrelevant AOIs. While it is not a recommended statistical practice to report both follow-up approaches to understanding an interaction between factors, here in the supplemental we share the alternative framing of the interaction follow-up (i.e. comparing across Position for each Task Relevance state) for information purposes only to illuminate the nuances of potential spatial biases. These spatial differences are secondary to the Relevant vs Irrelevant story, but may be informative for considerations to task design and the data processing method.
Therefore, to explore whether any spatial biases were present in our gaze data, we present the analysis of Relevant and Irrelevant gazes separately with two 4 x 1 (Position x Relevant/Irrelevant) RMANOVAs. Both single-factor RMANOVAs did reveal significant main effects of Position (Relevant: $F(1.54, 43.08) = 17.4$, $p < .001$; Irrelevant: $F(1.73, 48.34) = 20.4$, $p < .001$).

When AOIs are Relevant to a given movement, all AOI Positions receive similar lengths of looks except NRAOI, which receives significantly longer fixation time ($M = 0.817$ secs, all p’s $\leq .002$). Given the spatial proximity of NRAOI to the Home area (a non-AOI for our analysis purposes, but a “relevant” location during the task), this indicates that our clustering approach may have resulted in partially-inflated fixation durations for the NRAOI Position.

During moments when AOIs were Irrelevant to the task (i.e. not Pickup or Dropoff areas), pairwise comparisons showed significantly shorter gaze duration to the FLAOI than anywhere else ($M = 0.0767$ secs, all p’s $< .001$), and FRAOI exhibited the longest fixation duration ($M = 0.2791$ secs; significantly longer than NLAOI and FLAOI (both p’s $\leq .018$), but not NRAOI ($p = .393$)). The pattern of longer Irrelevant gazes to FRAOI indicate perhaps again some effects of nearby “non-AOI” spillover, where the starting fixation point, closest to FRAOI, was to be gazed at at the beginning of the trial. These brief fixation gazes may count as FRAOI-Irrelevant gazes given FRAOI’s non-involvement during the first movement of the trial. The shortest fixation durations towards FLAOI when FLAOI is Irrelevant may also reflect spatial spillover effects, but in the opposite way: FLAOI is the least likely to be gazed upon “in passing” on the way to a Relevant location as it’s the furthest from the repeatedly-returned-to Home area. Together, these results suggest some limited misattribution of non-AOI fixations as a function of the task’s properties (i.e. an AOIs spatial proximity to other “relevant” areas like Home and the fixation point) and the scope of the AOIs examined (i.e. not including other task-related AOIs). Most importantly, however, these results serve to confirm our eye-tracking method (both in the data collection and the data processing) as one that can be used to meaningfully explore more complex measures like EAL and ELL during online, screen-based, object interactions.

Phase Duration

Reach: Pairwise comparisons for the Reach phases show prolonged Reach phase durations for Moves 3 (longest), 5, and 7 compared to Moves 1,2,4,6 and 8 (all p’s $\leq .011$, except $2 > 7$, $p = .279$). Of the longer Reach phase movements, Move 3’s Reach phase ($M = 0.939$ secs) is also significantly longer ($p = .049$) than Move 7’s Reach phase ($M = 0.815$

secs).We can attribute a longer Reach of Move 3 than of Move 7 to a proximity to Home effect, where a reach from Home to the FLAOI (Move 3) is a further distance to travel than the Home-NLAOI reach, following Fitts’ law. Notably, Lavoie et al. (2018) also find, for their Reach (only) phase, a longer phase for Movement 3, citing the longest distance to traverse.

Transport: Move 8’s Transport phase is significantly shorter in duration than any other Transport (M = 0.143 secs; all p’s ≤ .01). Of the remaining movements, the Transport phases are relatively similar (M’s range from 0.205 to 0.255 secs), however Moves 3 and 7 are significantly shorter than Moves 5 and 6 (all p’s ≤ .047). The results suggest that Left-to-Right Transports are completed in a shorter duration, and that these Left-to-Right Transports get shorter as the trial continues, with the two final Left-to-Right Transports (and final movements - Moves 7 and 8) reflecting shorter and shorter durations.

Release: Pairwise comparisons between movements for the Release phase show significantly longer Release phase durations for Moves 2,5,6 and 7 (M’s range from 0.658 to 0.710 secs) compared to Moves 1, 3, 4, and 8 (M’s range from 0.441 to 0.570 secs; all p’s ≤ 0.045). Move 3 is also significantly faster than Moves 4 and 8 (both p’s ≤ 0.002).

Eye Arrival Latency (EAL)

Pick-up: The eye arrives at the Pick-up location earlier if the Reach phase is longer (e.g. Moves 3, 5 and 7), and less early (i.e. closer to the Pick-up time; shorter in preceding latency) if the Reach phase is shorter (e.g. Moves 1, 2, 4, 6, 8). Move 4’s EAL were the least early, significantly shorter than the EAL of all other movements (M = -0.337 secs, all p’s ≤ 0.047), and Moves 1, 2, 6 and 8’s EAL values were the next shortest (M’s range from -0.439 to -0.398 secs; no pairwise differences: all p’s > 0.438). Move 5 EAL was significantly earlier than all other movements (M = -0.535 secs, all p’s ≤ .041) except Move 3, the next earliest EAL (M = -0.482, p = .280). Move 7’s EAL value was also earlier (M = -0.457), and showed no difference in EAL to Move 3 (p = 1.00).

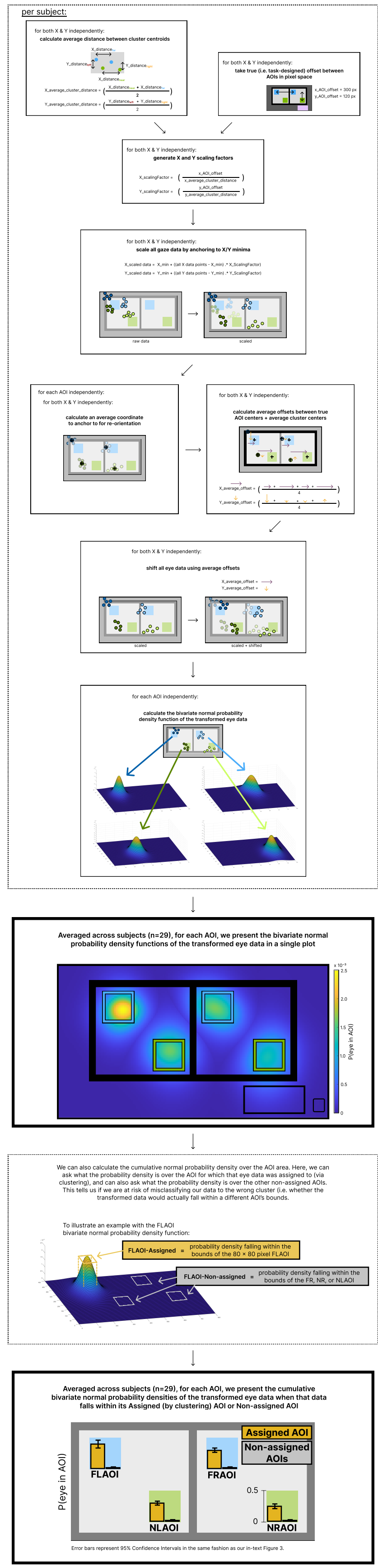
Eye Leaving Latency (ELL)

Pick-up: At Pickup, ELL remains relatively similar across movements (M’s ranging from 0.0362 to 0.1366 secs). Move 6 has the longest ELL (M = 0.1366), and is significantly longer than Moves 1 and 3 (both p’s ≤ .008). Move 1’s ELL at Pickup is the shortest (M = 0.0362), but besides the aforementioned Move 6, only Move 2 tested significantly longer than Move 1 (p = .025). These few significant differences represent a subtle push-pull effect of the duration of the Transport phase, and the preceding EAL values.

Drop-off: Moves 2, 5, 6, and 7 elicited the longest (and same; M’s range from 0.613 to 0.635 secs; all p’s = 1.00) ELL values, and were all significantly longer than any other moves (all p’s ≤ .018) except the pairwise comparison of Move 1 and Move 6 (p = .112). Next longest ELL at Drop-off was during Move 3 (M = 0.517 secs), which was significantly longer than Moves 4 and 8 (both p’s ≤ 0.013). EAL at Move 1 (M = 0.488 secs) showed no difference in EAL to Moves 3, 4 or 8 (all p’s = 1.00). These results follow the results of the Release phase, where longer Release phases show longer ELL values.

B.5 Spatial Assessment of Clustering Process

5.0 - Spatial Assessment of Clustering Process



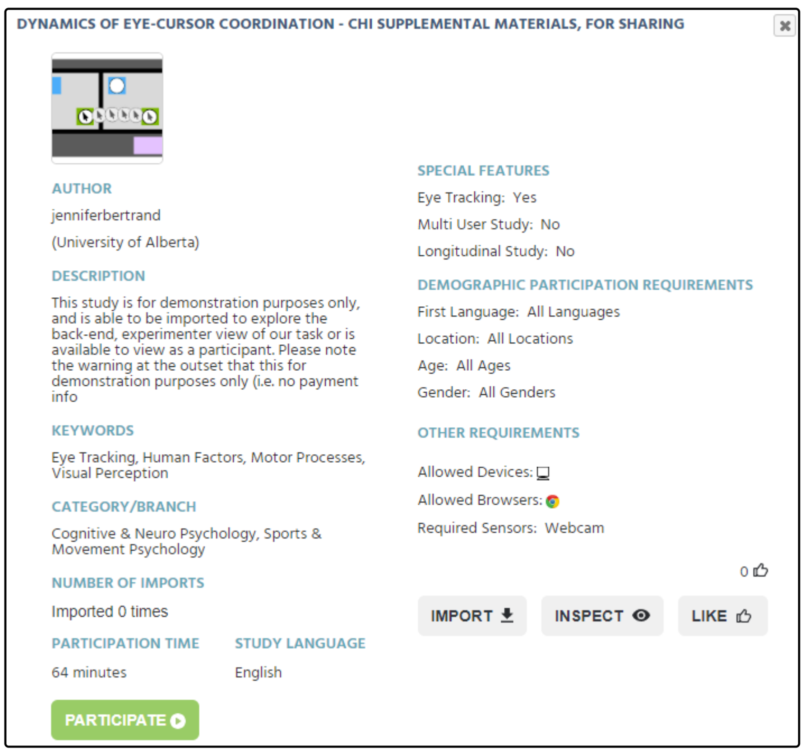
B.6 Labvanced Study Link

6.0 - Labvanced Study Link

The study can be accessed here: https://www.labvanced.com/page/library/39818

If the study link becomes broken, please use information from attached screen shot to search for thew study in the Labvanced library. Please contact the corresponding author if there's any issue with accessing the study.

The study is for demonstration purposes only, and is able to be imported into a free account or inspected to explore the back-end, experimenter view of our task, or, without creating an account, is available to view as a participant. Please note the warning at the outset that the participant view is for demonstration purposes only (i.e. no payment information applies, and no data will be used).



Appendix C: Supplemental Materials from Chapter 3 Pre-Print (Bertrand, Ouellette Zuk & Chapman, 2023)

Supplemental Materials 1 - Prolific Description

Choose between 2 presented options -with webcam-based eye-tracking (*need high-perf. comp, MOUSE ONLY, graphics card*)

By ualberta.ca

£6.50 • £6.00/hr | 65 mins | 30 places

Welcome to the experiment! This study should take approximately 65 minutes, and needs to be completed in one sitting. You will be required to stay **very still** during the entirety of the experiment for eye-tracking purposes.

Some people may not have the internet speed, graphics card, and/or hard drive system performance that this study requires - many factors contribute to this. **If you've closed other internet tabs etc. and you still get an error that says the face/video processing is not working well enough, you will not get past the very first screen. You will be unable to participate and will need to RETURN THE STUDY. PLEASE DO NOT KEEP TRYING - it will not work.**

A WIRED OR WIRELESS COMPUTER MOUSE IS ALSO REQUIRED. NO TRACKPADS/TOUCHPAD/TOUCHSCREEN/STYLUS. Do not try - it will not work for the experiment (please do not waste your time trying). **Thank you!**

We are conducting an academic study about decision-making. To study this, we need participants to choose between two presented options on the computer while a webcam-based eye-tracking system determines where the participants eyes are looking. While you complete this decision task, your eye gaze position will be recorded. While the webcam is used to calculate your eye gaze position in numerical form, no webcam video data is recorded.

Warning: some images shown may contain potentially explicit or offensive content (e.g. blood, violence) that some viewers may find disturbing. Participant discretion is advised.

We appreciate your interest in participating! This scientific data will be used to help us develop a better understanding of the eye-gaze in decision-making.
Still interested? Here's the important details:

Very important - please read the following requirements carefully

To participate, make sure you:
-- Use Chrome as your browser
-- Use a computer MOUSE (NOT A TRACKPAD OR LAPTOP TOUCHPAD ETC - WILL FAIL)
-- Use only a laptop or desktop computer (tablets/phones will fail)
-- Can do a computer task without needing to wear glasses (contacts OK)
-- Have a connected webcam (either built-in or external device is fine)
-- Have your laptop or desktop computer on a stable table top (not on your lap), in a well-lit room - with light in front of you (not behind) so that your face isn't shadowed. Cannot be in a dark room with only light coming from the computer screen (eye-tracking system will fail)
-- Are aged 18 to 35
-- Have the ability to keep a consistent, still head position for the full duration of the study
-- Can keep the study full screen for the full study duration (approx. 60 mins)
-- Close all other internet tabs before starting the study (experiment will fail)

The task will take approximately 60 minutes to complete. Complete the task in one sitting. You will not receive payment if you do not complete the task fully and in one sitting.

DO NOT exit from full screen mode during the experiment. If you are not using the Chrome browser, please swap browsers before beginning the study. The study will take 2-3 minutes to load at the beginning of the study.

VERY IMPORTANT: Please close all tabs before beginning the study. You will receive an error if your background activity is impeding the performance requirements of the eye-tracking system and you'll need to return the study if the performance threshold is not met at the beginning of the study (you will get an error screen that tells you this). Some people may not have the internet speed or system performance that this study requires. If you've closed other internet tabs etc. and you still get an error that says the face/video processing is not working well enough, you will not get past the very first screen. You will be unable to participate and will need to return the study - do not keep trying. Thank you.

This research is carried out by researchers from the University of Alberta, in Edmonton, Alberta, Canada. The plan for this research has been reviewed for its adherence to ethical guidelines and approved by Research Ethics Board 2 at the University of Alberta (Pro00087329).

Devices you can use to take this study: Desktop
You will also need: Camera

Supplemental Materials 2 - Labvanced Study Link

Experimental Task on Labvanced (for viewing purposes): https://www.labvanced.com/page/library/48965