

# A Representational Basis for Human-Computer Interaction

by

Barry Alan Po

B.Sc. (Honours), Queen's University at Kingston, 2001  
M.Sc., The University of British Columbia, 2002

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Computer Science)

THE UNIVERSITY OF BRITISH COLUMBIA

April 18, 2005

© Barry Alan Po, 2005



# Abstract

Mental representations form a useful theoretical framework for understanding the integration, separation and mediation of visual perception and motor action from a computational perspective. In the study of human-computer interaction (HCI), knowledge of mental representations could be used to improve the design and evaluation of graphical user interfaces (GUIs) and interactive systems. This thesis presents a representational approach to the study of user performance and shows how the use of mental representations for perception and action complements existing information processing frameworks in HCI. Three major representational theories are highlighted as evidence supporting this approach: (1) the phenomenon of stimulus-response compatibility is examined in relation to directional cursor cues for GUI interaction with mice, pointers, and pens; (2) the functional specialization of the upper and lower visual fields is explored with respect to mouse and touchscreen item selection; (3) the two-visual systems hypothesis is studied in the context of distal pointing and visual feedback for large-screen interaction. User interface design guidelines based on each of these representational themes are provided and the broader implications of a representational approach to HCI are discussed with reference to the design and evaluation of interfaces for time- and safety-critical systems, interaction with computer graphics, information visualization, and computer-supported cooperative work.



# Contents

<b>Abstract</b> . . . . .	iii
<b>Contents</b> . . . . .	v
<b>List of Tables</b> . . . . .	ix
<b>List of Figures</b> . . . . .	xi
<b>Acknowledgements</b> . . . . .	xv
<b>1 Introduction</b> . . . . .	1
1.1 Information Processing in HCI . . . . .	1
1.2 Why are Mental Representations Important? . . . . .	4
1.3 Scope and Objectives . . . . .	5
1.4 Overview of this Dissertation . . . . .	7
<b>2 Computational Perspectives on Human Vision</b> . . . . .	9
2.1 Computation and the Human Mind . . . . .	9
2.2 What are Mental Representations? . . . . .	11
2.3 A Three-Level Model of Computation . . . . .	12
2.3.1 Computational Description . . . . .	13
2.3.2 Representations and Algorithms . . . . .	14
2.3.3 Hardware Implementation . . . . .	15
2.4 Some Implications for Human Vision . . . . .	15
<b>3 Input and Interaction through Visual User Interfaces</b> . . . . .	19
3.1 Experimental Methods in HCI . . . . .	20
3.1.1 Psychophysical Techniques . . . . .	20

3.1.2	Fitts's Law and Related Models . . . . .	23
3.2	The Importance of Pointing as Interaction . . . . .	24
3.2.1	Mice and Relative Input Devices . . . . .	27
3.2.2	Pointers, Wands, and 3D Interaction . . . . .	28
3.2.3	Pens, Styli, and Touchscreens . . . . .	31
3.2.4	Voice and Multimodal Input . . . . .	33
3.3	Graphical Displays and Visual Output . . . . .	34
3.3.1	Monitors and Desktop Displays . . . . .	35
3.3.2	Large Screens and Immersive Displays . . . . .	36
3.3.3	Small Screens and Handheld Displays . . . . .	38
<b>4</b>	<b>A Representational Approach to HCI . . . . .</b>	<b>41</b>
4.1	Describing the Theoretical Framework . . . . .	41
4.2	Representations for Perception and Action . . . . .	43
4.2.1	Integration of Visual Information . . . . .	44
4.2.2	Separation of Visual Information . . . . .	45
4.2.3	Mediation of Visual Information . . . . .	45
4.3	Implications for HCI . . . . .	45
<b>5</b>	<b>Stimulus-Response Compatibility . . . . .</b>	<b>49</b>
5.1	Background and Related Work . . . . .	49
5.1.1	Previous Work in Ergonomics . . . . .	51
5.1.2	Previous Work in HCI . . . . .	53
5.2	Directional Compatibility and Cursors in GUIs . . . . .	53
5.2.1	A History of Pointing Cursors . . . . .	55
5.3	Hypotheses and Predictions . . . . .	58
5.4	Experiment 1: Comparing Cursor Orientations . . . . .	59
5.4.1	Participants . . . . .	59
5.4.2	Apparatus . . . . .	61
5.4.3	Procedure . . . . .	63
5.4.4	Data Analysis . . . . .	66
5.4.5	Movement Time Performance . . . . .	67
5.4.6	Precision and Accuracy . . . . .	70

---

5.4.7	Other Observations . . . . .	72
5.5	Discussion of Experiment 1 . . . . .	73
5.5.1	Implications for User Interface Design . . . . .	74
<b>6</b>	<b>Specialization of the Upper and Lower Visual Fields . . . . .</b>	<b>77</b>
6.1	Background and Related Work . . . . .	78
6.1.1	The Functionally Specialized Eye . . . . .	78
6.1.2	Reaction Time Performance . . . . .	81
6.1.3	Visual Attention . . . . .	82
6.1.4	Neuroanatomical Features . . . . .	83
6.2	Hypotheses and Predictions . . . . .	84
6.3	Experiment 2: Item Selection in the UVF and LVF . . . . .	85
6.3.1	Participants . . . . .	85
6.3.2	Apparatus . . . . .	85
6.3.3	Procedure . . . . .	88
6.3.4	Data Analysis . . . . .	90
6.3.5	Movement Time Performance . . . . .	91
6.3.6	Precision and Accuracy . . . . .	93
6.3.7	Overall Performance Across Input Methods . . . . .	95
6.3.8	Other Observations . . . . .	97
6.4	Discussion of Experiment 2 . . . . .	98
6.4.1	Implications for User Interface Design . . . . .	99
<b>7</b>	<b>The Two-Visual Systems Hypothesis . . . . .</b>	<b>101</b>
7.1	Background and Related Work . . . . .	102
7.1.1	Evidence from Neuroscience . . . . .	104
7.1.2	Evidence from Experimental Work . . . . .	106
7.2	Visual Illusions . . . . .	107
7.2.1	The Induced Roelofs Effect . . . . .	107
7.3	Hypotheses and Predictions . . . . .	110
7.4	Experiment 3: Visual Feedback on Large Screens . . . . .	112
7.4.1	Participants . . . . .	114
7.4.2	Apparatus . . . . .	114

---

7.4.3	Procedure . . . . .	116
7.4.4	Data Analysis . . . . .	117
7.4.5	Performance with Voice Input . . . . .	120
7.4.6	Pointing with Visual Feedback . . . . .	121
7.4.7	Pointing with Lagged Visual Feedback . . . . .	121
7.4.8	Pointing without Visual Feedback . . . . .	121
7.5	Discussion of Experiment 3 . . . . .	122
7.5.1	Implications for User Interface Design . . . . .	123
<b>8</b>	<b>Discussion and Applications . . . . .</b>	<b>125</b>
8.1	Theoretical Frameworks in HCI . . . . .	125
8.2	Developing New Interaction Techniques . . . . .	127
8.3	Ubiquity, Immersion, and Presence . . . . .	129
8.4	Application Domains . . . . .	131
8.4.1	Time- and Safety-Critical Systems . . . . .	132
8.4.2	Interaction with Computer Graphics . . . . .	134
8.4.3	Information Visualization . . . . .	135
8.4.4	Computer-Supported Cooperative Work . . . . .	137
8.5	Other Issues and Limitations . . . . .	138
8.6	Future Work . . . . .	141
<b>9</b>	<b>Conclusion . . . . .</b>	<b>145</b>
	<b>References . . . . .</b>	<b>149</b>
<b>A</b>	<b>Supplemental Data for Experiment 1 . . . . .</b>	<b>165</b>
<b>B</b>	<b>Supplemental Data for Experiment 2 . . . . .</b>	<b>179</b>
<b>C</b>	<b>Supplemental Data for Experiment 3 . . . . .</b>	<b>191</b>
<b>D</b>	<b>Glossary . . . . .</b>	<b>197</b>



# List of Tables

2.1	Three Levels of Complex Information Processing Systems . . . . .	13
2.2	A Representational Framework for Deriving Shape Information from Images	16
6.1	Summary of the Functional Differences in the UVF/LVF . . . . .	80
7.1	Individual Participant Performance in Experiment 3 . . . . .	119



# List of Figures

1.1	An Overview of the Model Human Processor in HCI . . . . .	2
3.1	A Collection of Relative-Mode Input Devices . . . . .	26
3.2	An Example of Laser Pointer Interaction with Large Screens . . . . .	29
3.3	An Illustration of Touchscreen Technology at Work . . . . .	32
3.4	A Desktop System with Multiple Monitors . . . . .	36
3.5	A User Working Inside a CAVE-Style Multi-Screen Display Environment . . . . .	37
3.6	A Selection of Small Screens and Handheld Displays . . . . .	39
4.1	Interactions Observed Between the Representations of a Visual Stimulus . . . . .	44
5.1	Pointing Cursors in WIMP-Based Graphical User Interfaces . . . . .	54
5.2	The Definition of a Cursor from the Xerox Bravo Reference Manual . . . . .	56
5.3	A Typical Trial Presentation in Experiment 1 . . . . .	60
5.4	Mouse Condition Apparatus Used in Experiment 1 . . . . .	61
5.5	Pointer Condition Apparatus Used in Experiment 1 . . . . .	62
5.6	Pen Condition Apparatus Used in Experiment 1 . . . . .	63
5.7	The Five Cursors Evaluated in Experiment 1 . . . . .	64
5.8	Bar Graph of Mouse Movement Times in Experiment 1 . . . . .	69
5.9	Bar Graph of Pointer Movement Times in Experiment 1 . . . . .	70
5.10	Bar Graph of Pen Movement Times in Experiment 1 . . . . .	71
6.1	An Illustration of the Upper and Lower Visual Fields . . . . .	79
6.2	A Typical Trial Presentation in Experiment 2 . . . . .	86
6.3	Experimental Apparatus Used in Experiment 2 . . . . .	87
6.4	Movement Time Regression Plots for Experiment 2 . . . . .	92
6.5	RMS Pointing Error Plots for Experiment 2 . . . . .	94

---

6.6	Bar Graphs of Overall Movement Time and Errors in Experiment 2 . . . . .	96
7.1	An Illustration of the Two-Visual Systems . . . . .	102
7.2	An Illustration of the Induced Roelofs Effect . . . . .	109
7.3	A Typical Trial Presentation in Experiment 3 . . . . .	113
7.4	The Experimental Apparatus Used in Experiment 3 . . . . .	115
7.5	Effect Size of the Induced Roelofs Effect in Experiment 3 . . . . .	120
A.1	Bar Graphs of Movement Time vs. Menu Position . . . . .	166
A.2	Bar Graphs of Movement Time vs. Menu Position . . . . .	167
A.3	Bar Graphs of Movement Time vs. Menu Position . . . . .	168
A.4	Bar Graphs of Movement Time vs. Menu Position . . . . .	169
A.5	Histograms of RMS Error Distribution . . . . .	170
A.6	Histograms of RMS Error Distribution . . . . .	171
A.7	Histograms of RMS Error Distribution . . . . .	172
A.8	Histograms of RMS Error Distribution . . . . .	173
A.9	Histograms of RMS Error Distribution . . . . .	174
A.10	Histograms of RMS Error Distribution . . . . .	175
A.11	Histograms of RMS Error Distribution . . . . .	176
A.12	Histograms of RMS Error Distribution . . . . .	177
A.13	Histograms of RMS Error Distribution . . . . .	178
B.1	Line Graph of Movement Time vs. Target Amplitude . . . . .	180
B.2	Line Graph of Movement Time vs. Target Width . . . . .	181
B.3	Line Graph of RMS Error vs. Target Amplitude . . . . .	182
B.4	Line Graph of RMS Error vs. Target Width . . . . .	183
B.5	Scatterplot of Mouse Pointing Errors by Target Amplitude . . . . .	184
B.6	Scatterplot of Touch Pointing Errors by Target Amplitude . . . . .	185
B.7	Scatterplot of Mouse Pointing Errors by Target Width . . . . .	186
B.8	Scatterplot of Touch Pointing Errors by Target Width . . . . .	187
B.9	Histograms of RMS Error Distribution for Mouse Pointing . . . . .	188
B.10	Histograms of RMS Error Distribution for Touch Pointing . . . . .	189
C.1	Histograms of Randomized Target Distribution . . . . .	192

---

C.2	Histograms of Randomized Target Distribution . . . . .	193
C.3	Histograms of Pointing Errors by Input Method . . . . .	194
C.4	Histograms of Pointing Errors by Input Method . . . . .	195
C.5	Bar Graph of Movement Times by Input Method . . . . .	196



# Acknowledgements

Although a single name appears as the sole author of a doctoral dissertation, anyone who has ever written one understands it is never quite possible without the support and assistance of others. I consider it a personal blessing that I have been around so many people who have been willing to help me make it this far. A piece of this dissertation (noted below or not) belongs to each of them.

To the members of my supervisory committee:

- Dr. Kellogg Booth has been my research supervisor over the last four years. I can honestly say that without him, I would have learned far less and that his approach to academia and research is one that will stay with me for the rest of my life. I will always be his student.
- Dr. Brian Fisher has been the strongest supporter of my research. His approach to HCI is what initially inspired all of the work reported here. Without him, this dissertation would not be what it is and I suspect my ability to fulfill the “prime directive” would have been substantially less gratifying.
- Dr. Ronald Rensink and Dr. Romeo Chua have always been supportive of my research, offering me guidance and support when I needed it. I hope they will be proud knowing their efforts have helped me succeed in an area of research where the end result is not always certain.

To the students, staff, and faculty of the Imager Lab for Computer Graphics, Visualization, and HCI:

- Especially Dr. Joanna McGrenere, Dr. Jason Harrison, Colin Swindells, Jennifer Fung, Allen Lin, Mark Hancock, and Alexander Stevenson, among others. Their research advice has always been valuable to me, as has their friendship. The Imager Lab in general has always been central to my ability to get the hardest work done and done well.

I would also like to thank the members of my examination committee:

- Dr. James Little and Dr. Peter Graf, who were my university examiners.
- Dr. Robert J. K. Jacobs from Tufts University, who was my external examiner.
- Dr. Matthew Yedlin, who was the chair of my examination committee.

The funding for the research presented here came from a variety of sources over the years, but most notably from the research and strategic grant programs of the Natural

Sciences and Engineering Research Council of Canada (NSERC). The University of British Columbia and its University Graduate Fellowship (UGF) program also provided financial support.

Finally, I would like to acknowledge the eternal support and patience given to me by (soon-to-be Dr.) Najwan Stephan-Tozy (soon-to-be D.M.D.), who has become the most important person in my life. Even in the hardest of times, she encouraged me to keep going and her understanding about what it means to be a graduate student has meant everything to me.



# Chapter 1

## Introduction

The sense of sight has had tremendous influence on the way that user interfaces and interactive systems are designed. It is difficult to imagine how much of the computing technology we use today would be possible without the ability to see. Because vision has always been so important, it is reasonable to argue that the study of human-computer interaction (HCI) has in many ways been the applied study of human vision as it relates to interface design. Even though most users have other senses such as audition and touch, the visual and graphical elements of a system are the ones that are most familiar to users and are probably the ones that many first identify as comprising the “user interface.” This has had a profound influence on the way that HCI is studied and practiced. The rapid emergence of graphical user interfaces (GUIs) and the movement toward direct, visually-guided interaction (i.e. direct manipulation) are strong examples of how the study of user interface design in HCI has predominantly been one of *visual* user interface design. Although the aspects of audition and touch remain important and are likely to be even more important in the future, the emphasis in this dissertation is on vision and the visual characteristics of user interfaces.

### 1.1 Information Processing in HCI

HCI has a history of studying interaction from the standpoint of information processing. Many of the empirical models of user performance that are now familiar in the discipline, such as Fitts’s Law and GOMS, are theoretically grounded from the perspective of users as information processors (Fitts, 1954; Card et al., 1983). The most prominent example of this approach in HCI was first described by Stuart Card, Thomas Moran, and Allen Newell (1983) in their classic description of the Model Human Processor, which brought together the results of earlier psychological research such as the work of Newell and Simon (1972), to build a model of routine cognitive tasks that could be used to understand performance in a variety of computing tasks involving humans. Their perspective has provided much of the

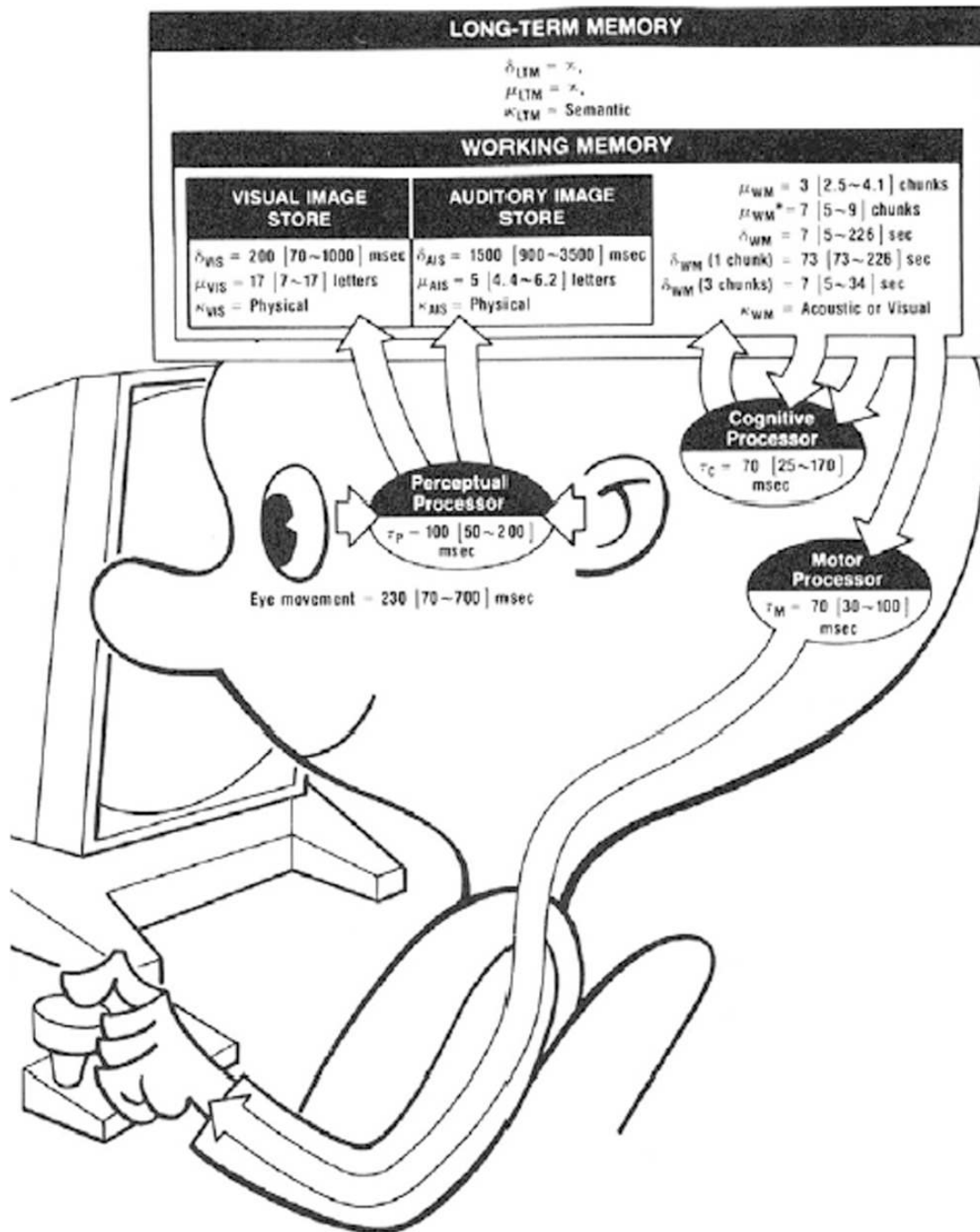


Figure 1.1: An overview of the Model Human Processor in HCI from Card, Moran, and Newell (1983). User performance is modeled as the amount of time spent with three interrelated subsystems: a perceptual subsystem, a cognitive subsystem, and a motor subsystem. In the Model Human Processor, an emphasis is placed on the relationship between visual perception and motor action and the internal representations that mediate them. Absent in this model are other sensory modalities like touch, although the model could be extended to include such components.

---

context for ongoing empirical work in HCI, especially where the optimization of particular user interface characteristics related to visual elements is concerned. Figure 1.1 presents an illustration of the Model Human Processor as originally described by Card, Moran, and Newell (1983).

In the Model Human Processor, user performance is attributed to resources spent on three interrelated subsystems: a *perceptual* subsystem that forms memories of detected sensory information, a *cognitive* subsystem that makes decisions based on these perceptual memories, and a *motor* subsystem that carries out physical movements and interactions with the interface. Implicit in this model of user behaviour is the belief that it is essential to characterize the *relationship* between vision and movement to understand user interaction, since neither vision nor movement constitutes interaction on its own. The human eyes form a part of the perceptual subsystem by acquiring images of detected light information. The processing of visual information by the human brain forms a part of the cognitive subsystem by allowing users to decide what the correct responses to visual events should be. The limbs are part of the motor subsystem, which enact the decisions made by the acquired and processed visual information.

The information processing approach taken by Card, Moran, and Newell has close ties to the modern study of cognitive psychology, which has been heavily influenced by computational theory and vice-versa. Of particular interest to many cognitive psychologists is the idea that human behaviour can be formally described as a collection of constructs known as *mental representations*. Such representations form the basis for understanding the integration, separation and mediation of visual information as it is used for *perception*, or input to the human visual system, and *action*, or actor-generated body movements (Prinz & Hommel, 2002). The importance of mental representations is echoed in the Model Human Processor but this is a topic that is largely unexplored in Card, Moran, and Newell. They wrote: “The perceptual system carries sensations of the physical world detected by the body’s sensor systems into *internal representations* of the mind by means of integrated sensory systems.” They indicate that something is intrinsically important about these representations to user performance, but this has yet to receive much attention in HCI. This thesis takes one step toward understanding the importance of mental representations and how they might be useful in describing and explaining the phenomenon of user performance.

## 1.2 Why are Mental Representations Important?

As user interfaces become more complex and interaction moves beyond the desktop, it will become necessary for HCI researchers and practitioners to understand more about vision, its relationship to movement, and the implications these have for the design of interactive systems. Though the technical aspects of hardware engineering and software implementation will always be important, such contributions are diminished without attention to how new technologies are processed by the perceptual, cognitive, and motor systems of the human user. As an example, Rensink (2002b) points out that developing “the conscious impression of a rich, coherent display requires the careful coordination of [visual] attention with the task at hand, so that external information can always be accessed when needed.” This is entirely independent of the hardware engineering or software implementation techniques used to implement the display.

In order to make human vision tractable and accessible to an audience concerned with the design of user interfaces, mental representations serve as a convenient model. A theoretical framework based on mental representations would allow HCI to look at the mechanisms of visual processing in a more intimate fashion, unifying the descriptions and models of user performance that already exist with the complexity of human physiology and the human brain. It provides HCI with a means for making new, and often counterintuitive predictions about user performance in various interactive settings. It also provides a means for evaluating the use of interactive technology in situations beyond those already envisioned, as is the case where input technology originally meant for use on the desktop is now finding use for large screen interaction and other kinds of interactive tasks.

In these ways, the study of mental representations can benefit HCI both theoretically and practically: a representational approach to HCI can help researchers gain a deeper understanding of the mechanisms driving user performance. It can help practitioners better understand the kinds of visual elements and organizational characteristics that constitute effective user interface designs, and it can be useful as a tool for justifying existing guidelines for user interface design and for developing new design guidelines. These benefits of investigating mental representations in HCI are not only limited to vision, but can also extend to how human users process information from other sensory modalities like audition and touch. Primarily for reasons of scope, this dissertation focuses specifically on mental

---

representations and the processing of visual information for movement.

### 1.3 Scope and Objectives

The central thesis of this dissertation is thus:

**A representational approach to the study of HCI is able to predict variations in visually-based user performance neglected by existing theoretical frameworks in HCI. Such insight can be helpful in the design and evaluation of interactive systems.**

A primary objective of this dissertation is to show how existing information processing frameworks in HCI (i.e. the Model Human Processor) can be augmented by an understanding of mental representations of visual space. Three different representational theories are examined to determine how user performance can be directly influenced by the ways in which visual information is internally represented by users. Our study of each representational theory presents a controlled empirical evaluation of a particular aspect of visual user interaction, demonstrating that models of the integration, separation, and mediation of visual representations for perception and action can systematically explain the performance characteristics of users in various interactive settings. A secondary contribution of this dissertation is the development of practical user interface design guidelines based on the investigation of these different theories, which provide an understanding of how mental representations can be helpful to the design of real-world systems. Specific applications to visual user interfaces for time- and safety-critical systems, interaction with computer graphics, information visualization, and computer-supported cooperative work (CSCW) are discussed as concrete examples.

The first major theory discussed in this dissertation examines the phenomenon of *stimulus-response compatibility*, which is already a topic of some interest in HCI (Fitts & Seeger, 1953; Worringham & Beringer, 1989; Chua, Weeks, Ricker, & Poon, 2001). Stimulus-response compatibility is argued to be a representation of perception and action that emphasizes the *integration* of visual information. This representational phenomenon is used to examine the issue of graphical cursor orientations with mouse, pointer, and pen input on small, medium, and large displays.

The second major theory discussed in this dissertation is the functional specialization of the *upper and lower visual fields*, which shows how separate representations govern the processing of objects in the visual world as a consequence of human evolution (Skrandies, 1987; Previc, 1990). The physiological differences between the visual fields are believed to contribute to a representation of perception and action that emphasizes the *separation* of visual information. This representational phenomenon is used to evaluate graphical item selection performance at different perceived locations on a large display using mouse or touchscreen input.

The third major theory discussed in this dissertation is the *two-visual systems hypothesis*, which states that different representations of visual information exist to enable the ability to complete perceptual tasks and to coordinate physical movement (Trevarthen, 1968; Milner & Goodale, 1995; Bridgeman, Peery, & Anand, 1997). The hypothesis suggests that the mental representations that integrate and separate visual information also mediate representations of visual information as they are used for perception and action. This representational phenomenon is used to motivate a study of how the presence or absence of graphical cues such as tracking cursors can influence user performance in large graphical display settings and immersive virtual environments.

Having explained what this dissertation is about, it is also important to say a few words about what this dissertation is *not* about. This dissertation presents an investigation of mental representations of visual space as it applies to HCI. It assumes that human vision can be described as a computational process. In cognitive psychology and other related fields such as kinesiology, this has been a topic of intense debate. The computational perspective appeals largely to a constructivist school of thought, and there are those, such as Searle (1997), who maintain that certain qualities of consciousness, such as intentionality and subjective quality, can never be computed. This thesis is not intended to decide which of these theoretical approaches are valid (or which ones are not). The question examined here is *not* whether the mind should be described in a particular way, but rather whether a particular perspective in cognitive psychology might be valuable to the study of HCI. The former is a matter for those in cognitive psychology and philosophy, while the latter is important to computer science and engineering. The approach presented here is largely a consequence of how valuable information processing models have been to the study of HCI and it is merely the intention of this thesis to show how such models can be augmented

---

and extended by a further understanding of the lower level mechanisms that underlie these descriptions.

## 1.4 Overview of this Dissertation

This dissertation has eight further chapters, broken down as follows:

- Chapter 2 presents the motivation and background of this thesis from the perspective of cognitive psychology. It summarizes the computational approach to understanding visual processing and how this relates to the representational theory of mind. A discussion of mental representations and how they relate to human comprehension of visual space is provided.
- Chapter 3 presents the motivation and background of this thesis from the perspective of HCI and interaction design. It reviews the relevant literature on input and interaction techniques and graphical display technology. Pointing is identified as a major interaction method and some discussion is provided regarding existing approaches to evaluating user performance in graphical pointing tasks.
- Chapter 4 ties together the material presented in the two previous chapters by describing how mental representations can influence user performance. Potential implications of this approach to user evaluation and the study of HCI are provided.
- Chapter 5 examines the first of the three major representational theories: stimulus-response compatibility. A description and literature review of related work are presented and an experiment comparing graphical cursors with implicit directional cues for interaction with small, medium, and large displays using mice, pointers, and pens is described.
- Chapter 6 investigates the second of the three major representational theories: the functional specialization of the upper and lower visual fields. A description and literature review of related work are presented and an experiment comparing mouse and touchscreen selection performance for items presented in the vertical hemifields is described.

- Chapter 7 looks at the third of the three major representational theories: the two-visual systems hypothesis. A description and literature review of related work with respect to the two-visual systems are presented. An experiment comparing voice input to pointing with varying degrees of visual feedback on a large screen graphical display is described
- Chapter 8 provides a general discussion of the representational approach, based on the evidence indicated by the study of the three representational theories. Specific comments about the application of this work to the study of HCI are presented with a special emphasis on the design and evaluation of interactive systems for time- and safety-critical situations, interaction with computer graphics, information visualization, and CSCW.
- Chapter 9 is the conclusion, which summarizes all of the work described in this dissertation and presents some closing remarks.

Several appendices are at the end of the dissertation. Appendices A, B, and C provide supplementary data for the three presented experiments that could be useful for future experimental replication or for further analysis of the experimental data. Appendix D provides a glossary of terms.



## Chapter 2

# Computational Perspectives on Human Vision

The study of human vision is one that stretches back to antiquity. Philosophers such as Plato and Aristotle addressed the question of what it meant “to see” long before modern scientific methods or experimental techniques even existed. This is a question that still remains mostly unanswered today, although the emergence of new experimental methods and new measurement technology has allowed vision scientists to make incredible discoveries over the last several decades. In the realm of cognitive psychology, new approaches to studying the sense of sight have provided tremendous insight into the internal mechanisms that permit people to see. Perhaps the most influential of these approaches to date is the *computational* theory of the human mind, which is very closely linked to the information processing models of user performance that have been so important to the study of HCI. Understanding the representational basis for HCI begins with an understanding of computation and its relationship to the understudy of human cognition.

### 2.1 Computation and the Human Mind

The computational theory of the human mind is one of the more popular schools of thought in cognitive psychology and is central to many of the current approaches for studying the nature of human behaviour. Computational theories arose during the 1950s when the dominant psychological approach was behaviourism, or the belief that the only psychological phenomena of interest were those that examined the relationship between observable stimuli and observable behavioural responses (Watson, 1913; Skinner, 1953). There was very little interest in studying the nature of consciousness or the human mind during the reign of behaviourism because these were considered to be outside the realm of appropriate scientific study.

When the psychologist George Miller summarized the results of several studies demonstrating a limited capacity for human short-term memory of about seven items, he proposed that short-term memory actually required *mental representations* for encoding and decoding information (Miller, 1956). Miller's work happened at around the same time that other researchers, including Allen Newell (who was one of the key architects of the Model Human Processor described in the last chapter), John McCarthy, Marvin Minsky, Noam Chomsky, and Herbert Simon, were founding the study of cognitive science (Newell & Simon, 1972; Thagard, 1996). In the 1950s, theories of computation were gaining substantial interest, particularly with respect to the study of artificial intelligence and because of this, many had begun to question whether it was possible to reconcile the physical nature of the human brain with the immaterial nature of the human mind using computational models.

The central argument of the computational theory of mind is that thinking can best be understood in terms of representational structures and the computational procedures that operate on these structures (Thagard, 1996). These representational structures are analogous to data structures and the computational procedures that operate on them are analogous to formal algorithms. Thus, this approach to studying the human mind has direct links to theories of computation as they are studied in computer science and it is possible to see many aspects of one in the other. The key challenge of the computational theory of the human mind lies in understanding the *nature* of the representations and computations that constitute thinking. Disagreements arise in cognitive psychology and philosophy because there is no consensus on the computational approach that will eventually lead to a complete theory of the mind. Current theories about the kinds of computations that occur in the human mind include formal logic, rules, concepts, analogies, neural networks, and mental imagery (Chomsky, 1965; Pylyshyn, 1984; Kosslyn, 1987; Pinker & Mehler, 1988).

However, HCI is less concerned with this mind-body problem and is more interested in all of the various ways in which the mental representations that are hypothesized to operate in the mind could support an understanding of user interaction. Although this dissertation focuses on computation and representation as they pertain to visual perception and action, the many other ways in which representations model behaviour could also play an important role in HCI research. Sasse (1997) has argued that mental representations could be important in solidifying the concept of mental models in HCI, especially with respect to the ideas surrounding conceptual design.

---

It is perhaps unsurprising that user modeling in HCI has been heavily influenced by theories of computation and that information processing models of user performance still remain important. Many of the earliest attempts to formally model user performance were conducted by psychologists who believed in the potential of computational models to explain human behaviour. This should at least suggest to those studying HCI that understanding the mental representations that underlie a computational model may be as important as the models themselves.

## 2.2 What are Mental Representations?

Because mental representations are central to the computational theory of mind (and this dissertation), it is important to define the term carefully. Although this dissertation exclusively focuses on representations as they pertain to human vision and visual processing, such representations also serve as convenient models for many other aspects of human behaviour, including language, reasoning, and deduction. The use of mental representations for explaining behaviour actually predates the computational theory of mind, going at least as far back as Aristotle, who claimed that commonsense mental “states” such as thoughts, beliefs, desires, perceptions, and images were representations themselves (Cohen, Curd, & Reeves, 2000).

Johnson-Laird (1983) has suggested that mental representations are a means to reify the abstract concept of a mental model. He proposed that reasoning about a problem can be facilitated if a person has a mental model that represents the relevant information in an appropriate fashion for the problem to be solved. Others, such as Paivio (1986) argue that mental representations in fact have an analogy to physical representations, leading to a definition of representation that spans a continuum from picture-like representations to language-like representations. Thus, depending on whose computational theory of mind is being discussed, what is actually meant by a “mental representation” can change quite substantially.

Among computational theorists in cognitive psychology, perhaps the only consistent property of a mental representation is that it is an abstract, information-bearing construct of some kind, though many would argue that mental representations are systematic structures as well (Fodor, 1983; Pylyshyn, 1984). Because of this and because there are so many

different kinds of hypothesized mental representations, not all of which are relevant to this dissertation, we will define a mental representation as follows:

**Mental Representation** An abstract structure that encapsulates information about external events in a systematic way so that appropriate decisions may be made.

This definition suitably encompasses all of the relevant computational approaches to studying vision and the mind without loss of generality to other bodies of work that use the term. Furthermore, it is a definition that is sufficiently simple and accessible to the study of user performance in HCI, which is especially important in a discipline that must meet the needs of both researchers who are particularly interested in theory and practitioners who are particularly interested in building systems. It is a definition that also emphasizes the scope of this dissertation, which is focused on showing how different kinds of mental representations can influence user performance and not on the particulars of syntactic properties, physical realization, or their philosophical implications.

### 2.3 A Three-Level Model of Computation

Perhaps the best studied component of the “computational mind” is the human visual system. Much of the empirical evidence supporting a computational theory of the mind is actually derived from theoretical predictions of how various representations of visual information induce specific patterns of behaviour. It is also an approach to studying vision that has been profoundly influenced by the cognitive psychologist David Marr (Marr, 1982). Marr in particular believed that it was important to reconcile the “hardware” of the human visual system (the brain) with its less tangible aspect of enabling many kinds of behaviour. His research has left a lasting impression on the way that the particular problems posed by vision are understood and his theories on information processing have had an impact on computational models of the human mind in general.

Marr believed that human vision needed to be modeled as a computational system because complex information processing systems were always more than the sum of their parts. Such systems could never be understood as a simple extrapolation of their elementary components. He proposed that any complete theory of a complex system, such as a theory of human vision, needed to address three fundamental issues. First, the theory needed to

1. Computational Description	2. Representation and Algorithm	3. Hardware Implementation
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?	How can this computational description be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?	How can the representation and algorithm be realized physically?

Table 2.1: The three levels of complex information processing systems, derived from Marr (1982). These were the levels at which any machine carrying out an information processing task needed to be understood.

enumerate *what* the end goals of the system were. Second, the theory needed to explain *why* such goals were necessary or appropriate. Third, the theory needed to describe *how* a system constructed for the purpose of accomplishing these goals could do so.

According to Marr, these three fundamental issues could be resolved by defining an information processing system like human vision in terms of three levels of abstraction. Such systems contain a level of *computational description*, a level of *representation and algorithm*, and a level of *hardware implementation*. In his model, all three levels are logically and causally related, meaning that they could be interpreted as forming an implicit hierarchy of abstraction, with computational descriptions forming a “high-level” description of the system and hardware implementations forming a “low-level” description. Table 2.1 summarizes Marr’s three-level information processing model.

### 2.3.1 Computational Description

Computational descriptions form the high-level description of an information processing system. Marr gives “description” a very specific definition at this level: it is the result of using a representation (defined at the mid-level of the system) to describe a given entity (Marr & Nishihara, 1978). It is perhaps somewhat unfortunate that this is called a “computational” description, since it is important to realize that his entire model is a computational abstraction, and not just this particular level. Nevertheless, this is the level at which the *performance* of the system is characterized as a mapping from one kind of information to another. This is often called the *functional* description of the system. An important prop-

erty of the computational description is that the abstract properties of this mapping are precisely defined. Another important part of the computational description involves the justification of appropriateness and adequacy: why has the system been designed in this particular way and how has this served to fulfill its functional goals?

Marr believed that the primary goal of the human visual system was to construct a three-dimensional representation of distal stimuli on the basis of inputs to the retina that served to promote useful behaviour with the surrounding environment. Useful behaviour in this sense meant that vision had a wide variety of functions, including the ability to identify shapes and objects at a distance and the ability to support navigation and movement. In this sense, his interpretation of vision coincided with the ecological approach taken by Gibson (1979), who regarded the “problem” of vision as being that of utilizing the valid properties of the external world from visual information, although Marr believed that Gibson seriously underestimated the complexity of the information processing problems posed by vision.

### 2.3.2 Representations and Algorithms

Representations and algorithms form the mid-level description of Marr’s information processing model. At this level, the choice of representation for the input and output is described as well as the algorithm that is used to transform the input into the output. Marr chooses to define a representation as a formal system for making explicit certain entities, or types of information, together with a specification of how the system does this. This particular definition has the critical implication that information represented in one way can facilitate the particular goals of a system while the very same information, when represented in another way, may not. Marr refers to the different ways in which numbers can be represented as an example: it is easy to add, subtract, or multiply when numbers use an Arabic or binary representation, but it is not at all easy to perform these same operations as Roman numerals.

Instead of vision being the computational result of abstract “invariants,” or permanent properties of the environment as Gibson believed, Marr thought of vision as being no more complicated than mapping from one representation to another. One of the major implications for this representational approach versus Gibson’s ecological approach was that the processes of vision could be specified in very precise terms without deferring to a sensory apparatus of arbitrary complexity. Thus, one could imagine that visual information, at least

---

in its initial representation, could be defined as multiple arrays of image intensity values as detected by the photoreceptors in the retina, which would in turn be algorithmically transformed into a representation useful for a particular visual task, such as pointing at an object in the surrounding world.

### **2.3.3 Hardware Implementation**

Hardware implementations form the lowest-level description of Marr's information processing system model. It is analogous to the detailed architecture of the system, which describes how the system is physically "implemented." This level of detail is important because the particular way in which a representational abstraction is implemented can have consequences for the performance of the representation itself.

The "hardware" of human vision includes the neural connections and cells that permit the transduction of light information detected by retinal cells. When the level of hardware implementation is combined with that of the levels of computational description and representation, one begins to see how it is possible to reconcile the apparent seamlessness of the visual experience with the complex neural hardware that apparently makes such experience possible. Marr points out that one could argue against an information processing model of vision by appealing to the presence of visual illusions and other perceptual ambiguities as philosophers have done in the past (Austin, 1962). This particular argument can be countered when one considers how visual information is implemented at the level of the human brain. One possible consequence of human evolution is that the human visual system has been implemented as a system capable of performing a great many things, but perhaps at the expense of other functions that were less appropriate to ancestral survival. Thus, the existence of visual illusions is not evidence that vision is something other than computation but rather evidence that vision may not have been implemented to deal with these particular kinds of visual phenomena.

## **2.4 Some Implications for Human Vision**

The three-level information processing model introduced by Marr led him to believe that a representational framework for vision divided the derivation of shape information from images into three representational stages: (1) the representation of properties of the 2D

image such as intensity changes and local 2D geometry, (2) the representation of properties of the visible surfaces such as distance from the viewer and surface discontinuity in a viewer-centered coordinate system, and (3) the representation of the 3D structure and organization of the viewed shape from an object-centered perspective. Table 2.2 summarizes Marr's basic framework for human vision.

<b>Name</b>	<b>Purpose</b>	<b>Primitives</b>
Image	Represents intensity	Intensity value at each point in the image
Primal sketch	Makes explicit important information about the 2D image, primarily the intensity changes there and their geometrical distribution and organization	Zero-crossings, blobs, terminations and discontinuities, edge segments, virtual lines, groups, curvilinear organization, and boundaries
$2\frac{1}{2}$ D sketch	Makes explicit the orientation and rough depth of the visible surfaces, and contours of discontinuities in these quantities in a viewer-centered coordinate frame	Local surface orientation (the "needles" primitive), distance from viewer, discontinuities in depth, discontinuities in surface orientation
3D model representation	Describes shapes and their spatial organization in an object-centered coordinate frame, using a modular hierarchical representation that includes volumetric primitives (i.e. primitives that represent the volume of space that a shape occupies) as well as surface primitives	3D models arranged hierarchically, each one based on a spatial configuration of a few sticks or axes, to which volumetric or surface shape primitives are applied

Table 2.2: Marr's (1982) basic representational framework for human vision, based on deriving shape information from images. He argues that processing visual information, such as the shape properties of an object, is the result of multiple representational transformations.

Although much of Marr's representational framework has been supplanted by more recent empirical evidence (i.e. his concept of a unified percept of the visual world has been challenged by the work of Milner and Goodale (1995) and others), the three-level information processing model continues to have value in the study of visual behaviour. For example, representations of visual information have been important to understanding phe-



---

nomena such as visual attention, which have traditionally eluded systematic study. Moreover, the concept of representations in visual processing has been influential in structuring visual processing in terms of computational modules, analogous to the modular organization of the mind proposed by Jerry Fodor (1983). This architectural model of vision has also served to support the study of visual processing as a “vertical” structure, with simultaneous top-down influences such as expectations, memories, and biases, and bottom-up influences such as direct changes to the visual world itself that can affect visual perception (Long & Toppino, 1994; Cave & Kim, 1999; Itti & Koch, 2000).

Advances in the neuroscience and physiology of vision are directly linked to specific perceptual effects such as change blindness, or the inability to detect a change in a visual scene under certain circumstances despite the fact that such changes would normally be easy to detect (Simons & Levin, 1997; Rensink, 2000, 2002a). Other research has shown that individual visual items become “lost” in the presence of peripheral distractors, suggesting that there are limits to the attentional resolution of the human visual system (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Cavanagh, 2001). This limitation appears to be supported by research showing that the human visual system may also be bounded by limited computational resources. The FINST (“FINgers of INSTantiation”) theory proposed by Pylyshyn (1994; 2003) suggests that visual indexing occurs, allowing humans to keep track of between four and six visual objects at a time. The neural and behavioural correlates of these and other effects give rise to interpretations that can be used to better understand the specific characteristics and limitations of visual performance.

Research into human vision is ongoing and it will likely be many more years before anyone can offer a complete theory of vision that ultimately fulfills the requirements of Marr’s three-level information processing model. Though the research presented above does not give a complete impression of all the different kinds of vision research currently taking place, it is readily apparent that the notion of representations and the computational approach might offer insights into human visual processing that might have implications for HCI and user interface design. As we will see in the coming chapters, the fact that the same detected visual information can be represented in multiple ways to achieve different goals means that these multiple, simultaneous representations of vision can have a systematic, observable impact on user performance.



## Chapter 3

# Input and Interaction through Visual User Interfaces

The design and evaluation of input and interaction techniques is central to the study of HCI. Because user interfaces and technology have changed so much in a very short period of time, it is important to review some of the research that has occurred since the late 1970s. When user interfaces were first being evaluated, desktop computers were not yet commonplace and much of the research had an emphasis on the ergonomics, or human factors, of systems that enabled repetitive tasks such as textual transcription and telephone operator tasks (Card et al., 1983). Many of the lessons learned in these early user systems continue to be applied today, but in a considerably different context.

Computers and the user interfaces that go along with them are now seeing use in more ubiquitous scenarios. Mobile phones and personal digital assistants (PDAs) are widely used throughout the world and large screen displays seem to be used in conjunction with portable computers to form ad-hoc collaborative environments. Touchscreens have become a viable method for user interaction and computer vision is rapidly enabling new forms of input, such as 3D gesture. These and other new applications for interaction technology point to a greater need for design and evaluation methods that augment existing techniques in HCI. There is a tighter coupling between visual perception and motor action today than there has ever been in the history of user interface design and the gap between our understanding of user performance and the technology that enables interaction continues to widen. A representational basis for perception and action may be integral to closing this gap by providing a theoretical grounding for understanding the usable characteristics of existing and future technology.

## 3.1 Experimental Methods in HCI

Existing evaluation methods in HCI are derived from a number of disciplines, including engineering, cognitive psychology, and the social sciences. Techniques can be grouped into one of three categories: user modeling (i.e. GOMS and keystroke-level models), qualitative methods (i.e. ethnography and cognitive walkthroughs), and quantitative methods (i.e. empirical studies and user preference questionnaires based on the Likert scale). In this dissertation, we borrow from the quantitative group of evaluation techniques and use controlled experiments to study the impact of representations of visual space. The experiments discussed here are largely influenced by the *psychophysical* method, which is a reliable technique for studying certain kinds of behavioural phenomena in cognitive psychology and neuroscience.

### 3.1.1 Psychophysical Techniques

Psychophysical experimentation dates back to the nineteenth century when the study of psychology was still in its infancy and there were no existing physiological methods that enabled the objective evaluation of sensory or neural functions. Gustav Fechner is generally credited with the development of psychophysics as a means of studying the relationship between the mind and body (Adler, 1966). Fechner believed that mind and body were just different reflections of the same reality, which is a belief that resonates with the computational theory of mind in cognitive psychology. To study these “reflections,” Fechner devised methods for establishing a correlation between neuronal (i.e. objective) and perceptual (i.e. subjective) events.

The necessity for psychophysical techniques arises because there is an apparent conflict between obtaining perceptual experiences, which everyone has, and investigating them so they can be communicated and shared by others. In modern psychology, psychophysical techniques are used to selectively isolate given neural mechanisms by the behavioural patterns they are hypothesized to generate. Experimenters infer the existence of a particular neural mechanism for behaviour based on objective observations using physical or virtual stimuli (Wist, Ehrenstein, & Schrauf, 1998). Thus, physical or virtual stimuli are used as a reference for particular stimulus characteristics that are systematically manipulated to measure observer performance.

---

Four basic behavioural tasks are assessed using psychophysical techniques: detection, identification, discrimination, and scaling (Ehrenstein & Ehrenstein, 1999). Detection tasks are arguably the most basic, focusing simply on whether an observer is able to perceive a stimulus at all. Identification tasks build on detection by further asking observers the specific characteristics of a stimulus they can identify, such as spatial location. Discrimination tasks ask observers to detect or identify stimuli under uncertain conditions, including situations where environmental noise might mask the presence of a relatively weak stimulus. Scaling tasks measure the magnitude of stimuli by systematically measuring observer responses on a psychophysical scale.

There are particular kinds of psychophysical tasks and methods that have been valuable in studying behavioural phenomena. There are methods based on threshold measurements, which are used to identify the intensity at which observers can just barely detect a stimulus (sometimes known as the absolute threshold) and to identify the minimum intensity at which a variable comparison stimulus must deviate from a constant control stimulus to produce a noticeable perceptual difference (sometimes known as the difference threshold or the just-noticeable difference). There are forced-choice methods, which require observers to make a response on every trial in an experiment regardless of whether the observer “perceived” the presentation stimulus; such methods can be used to reveal detection or identification thresholds beneath a hypothesized absolute threshold. There is also the signal detection approach, which is used to concurrently measure the sensitivity of observers in performing a perceptual task and any response bias they might have.

In HCI research, variations on psychophysical techniques are used to conduct experiments on tasks that require visually-guided motor movement. Basic visual stimuli are used in target acquisition tasks where participants are required to engage in physical movement to make a response in individual trials. Trials are commonly blocked according to the experimental factors that are being varied and these trials are typically sufficient in number so that each participant can be considered to be an independent experiment (Vicente & Torenvliet, 2000). Many experiments employ completely within-subject designs and individual trial combinations are repeated several times to minimize response variance. These kinds of experiments can be used to identify systematic variations in user performance based on variations in the interactive environment (i.e. different input devices, graphical displays, or interface type) and have been useful in characterizing the limits of user performance,

building general design guidelines for user interfaces, and directly applying the results of psychological experiments to the design of interactive systems (Ware, 2004). Many of the key empirical results that were used to build the Model Human Processor described by Card, Moran, and Newell (1983) are based on experimental data collected using these kinds of techniques.

### Quantitative Metrics

Motor performance experiments often focus on two particular kinds of quantitative metrics: response time and accuracy. Response time could specifically be any number of measures, including *response delay* defined as the initial onset of movement from the presentation of a visual stimulus, *movement time* defined as the time between initial onset of movement and the termination of the movement response, or frequently it is the sum of both response delay and movement time. Although in some cases, experiments yield response time differences across experimental conditions that can be measured in seconds, response time differences are often measured in milliseconds because the hypothesized differences in performance are only visible at this level of resolution. In experiments where effect sizes might be especially small (perhaps less than 50 milliseconds), high-resolution timers, independent timing devices, and real-time operating system environments must be employed to provide some guarantees as to the accuracy of the timing measurements.

Similar to response time, the measures for motor accuracy depend on the phenomena being examined. Accuracy could be defined with respect to target position, such as how far a response was from the center of a trial target, for which a better term may actually be motor precision. In some experiments, accuracy could be restricted to a particular axis of movement (horizontal or vertical) to simplify the data analysis. In other cases, accuracy may not be a continuous measure at all. Instead, accuracy may simply be binary (“Did the participant hit the target?”). In kinesiology and cognitive psychology, it is common to see accuracy measured in terms of degrees of visual angle, or the number of degrees subtended by the viewer’s eye at a given distance. This is also seen in HCI, especially for the more theoretical modeling experiments, but it is also common to see accuracy measured in terms of pixel distance relative to the effective graphical resolution of a particular display.

Researchers conducting motor performance experiments and those interested in HCI often share similar quantitative metrics. However, their reasons for using such measures may

differ substantially. In disciplines like cognitive psychology and kinesiology, differences in quantitative measures may be used to infer the engagement of particular neural mechanisms related to vision and movement. In HCI, such measures are used in a comparative sense to learn what interface features facilitate improved user performance. This may be because of the traditional emphasis in user interface design for completing repetitive user tasks as efficiently as possible. Although such an emphasis is potentially less important in the context of a less tangible goal like enhancing the “user experience,” measures such as response time and motor precision continue to be important in quantitatively assessing usability.

### 3.1.2 Fitts’s Law and Related Models

An entire experimental theme in HCI is devoted to the study of pointing performance models based on Fitts’s Law (Fitts, 1954). This is an empirical model that is touched upon in the discussion of stimulus-response compatibility (Chapter 5) and is explicitly applied in the investigation of the upper and lower visual fields (Chapter 6). Fitts’s Law is a model of visually-guided target acquisition sometimes explained using an approximation of Shannon’s Theorem in information theory (Shannon & Weaver, 1949; MacKenzie, 1989). Fitts’s Law defines a log-linear relationship between movement time, target size, and target distance, which is commonly expressed as the following equation:

$$MT = a + b \log_2\left(\frac{A}{W} + 1\right) \quad (3.1)$$

where  $MT$  refers to movement time,  $A$  refers to movement amplitude or target distance, and  $W$  refers to target width. The constants  $a$  and  $b$  are usually derived empirically and can be interpreted as the  $y$ -intercept and slope of a predictive linear regression equation. The overall expression  $\log_2\left(\frac{A}{W} + 1\right)$  is frequently referred to as the *index of difficulty*, which indicates how difficult a presented target is to hit. There are several formulations of Fitts’s Law, including the original version presented by Fitts (1954), an alternative formulation presented by Welford (1960), and an information theoretic derivation by MacKenzie (1989). The particular formulation shown in Equation 3.1 is the MacKenzie formulation and is the one in common use in HCI. This is also the formulation used in this dissertation.

Regardless of the specific formulation, Fitts’s Law predicts that there will be a linear increase in movement time with a linear increase in the calculated index of difficulty. The

general implication of this prediction is that smaller, distant targets will be more difficult and will take more time to acquire than larger, nearby targets. The original task used to validate Fitts's Law involved reciprocal tapping by moving a stylus between two metal bars as quickly as possible. Even though this is not strictly representative of a task users might accomplish on a computer, the basic predictions suggested by Fitts's Law have proven experimentally to be extremely robust. Fitts's Law has been used to model movement performance under many different conditions where empirical data match the predictions of the model. These include extensive evaluation of mouse pointing, the effects of hand lag in virtual displays, modeling the movement and distribution of hits on a virtual keyboard, and the quantitative analysis of scrolling techniques (Card, English, & Burr, 1978; Boritz, Booth, & Cowan, 1991; MacKenzie, Sellen, & Buxton, 1991; Ware & Balakrishnan, 1994; Zhai, Sue, & Accot, 2002; Hinckley, Cutrell, Bathiche, & Muss, 2002).

Because Fitts's Law only refers to univariate movement along a horizontal axis of movement, subsequent work has modified the model to accommodate bivariate movements along arbitrary directions of movement and to allow for variations in the index of difficulty calculation so that additional parameters other than target width are taken into account (MacKenzie & Buxton, 1992; Accot & Zhai, 2003). Accot and Zhai (1997, 1999) have developed trajectory-based models based on Fitts's Law to analyze tasks such as drawing and writing, and Guiard, Beaudoin-Lafon, and Mottet (1999) have shown that the Fitts pointing model can be extended to navigational movements as a form of multi-scale pointing.

## 3.2 The Importance of Pointing as Interaction

A recurring theme in this dissertation is that of pointing. Pointing serves as a model for many kinds of user interaction and is also a task that happens to be quite well studied in the psychological literature. As a means for nonverbal communication, pointing has been a prominent example of the semantic and cultural implications of gesture in humans (Ekman & Friesen, 1969; Brinck, 2004). From this semiotic (linguistic) perspective, pointing is a particularly challenging gesture to categorize and understand because there are many syntactic nuances that are difficult to model. Kendon (1996) suggests that pointing has at least four different meanings. Pointing can be used to refer to physical objects that surround participants (actual object pointing), to objects that can have a physical location but are



---

not immediately present (removed object pointing), to objects that are given locations for the purposes of the current discourse (virtual object pointing), and to things that cannot have any sort of object status (metaphorical object pointing).

In the context of HCI, pointing has traditionally referred to the first of these definitions (actual object pointing) and pointing has become synonymous with acquiring targets on a graphical display. Even in this “limited” context, the exact meaning of pointing as interaction depends greatly on the interactive setting. Common experience tells us that pointing with a mouse is inherently different from pointing with a pen, so it is perhaps better to consider the act of pointing as encompassing an entire class of user interactions instead of thinking of pointing as serving a particular functional goal. Pointing has roots in HCI that can be traced back to the Memex envisioned by Vannevar Bush (1945), the idea of Man-Computer Symbiosis described by J. C. R. Licklider (1960), the oNLine System (NLS) of Douglas Engelbart (1963), and the conceptual Dynabook system of Alan Kay (1977). The popularity of the paradigm of direct manipulation moved pointing into the mainstream as a means of interaction, and the “Put-That-There” demonstration of Richard Bolt (1980) continues to serve as an example of how pointing might fit into the world of multimodal interaction (Shneiderman, 1982, 1983).

Various taxonomies have been proposed to categorize input devices for pointing and user interaction. The standards derived from the ACM Core Graphics System (1977) proposed a distinction between the physical structure of a device and its logical function. In this framework, devices were assigned to logical categories based on how they performed user actions. Buxton (1986) organized input devices according to properties of position, motion, and pressure, and the degrees of freedom afforded by the device. A semantic approach to categorization was taken by Mackinlay, Card, and Robertson (1990), who organized input devices according to their expressiveness and effectiveness. Jacob, Sibert, McFarlane, and Mullen (1994) took a perceptual approach to categorization, extending the theory of processing perceptual structure to the control structure of input devices, emphasizing the integrality and separability of perceptual attributes to show how the perceptual and control structures of a task map onto one another. Another approach taken by Guiard et al. (2004) suggests that pointing interactions can be decomposed into categories independent of input device. In their taxonomy, pointing interactions are categorized according to user view: cursor pointing, “Prince” pointing such that some interval must be moved to eventually

include a target, and view pointing such that the view itself must be moved in order to acquire a target.

### 3.2.1 Mice and Relative Input Devices

The mouse is the most popular of all pointing devices available today and is arguably the most representative of devices that fall into the category of *relative-mode* input. Figure 3.1 presents a variety of different relative-mode input devices. Relative input devices use changes in motion to determine the current location of user input and have no absolute origin, reporting only changes from their former position (Foley, van Dam, Feiner, & Hughes, 1997). Other common relative-mode input devices include trackballs and joysticks, which are more common in vehicle operation and industrial settings. The invention of the mouse is credited to Douglas Engelbart (1963) but has undergone substantial physical changes since its conception. Many of these changes are cosmetic in nature but some are functional, including the use of wireless mice and optical technology instead of moving mechanical parts to track user input.

Perhaps more substantial than the particular physical or hardware changes to these input devices are the ways in which modern GUIs and interaction design in HCI have emphasized the use of these devices as a form of “indirect” pointing. When Engelbart first envisioned the use of the mouse he foresaw its use as part of an efficient means for two-handed interaction with a computer different from the way in which the mouse is used today. Today, mice are the primary means for interacting with GUIs that adhere to the Windows-Icons-Menus-Pointer (WIMP) metaphor. When mice or other relative-mode pointing devices are used, visual feedback is almost always provided in the form of a tracking crosshair or arrowhead cursor.

User performance with relative input devices is well-studied in HCI. Among the first empirical evaluations was a study by English, Engelbart, and Berman (1967) who were interested in determining the best display selection techniques for computer-aided text manipulation. Card, English, and Burr (1978) conducted a subsequent study comparing a different set of input devices. In these early studies, it became clear that different input devices conferred particular advantages and disadvantages to a user because of the differences in their physical designs. Mice, in particular, tended to fare well when compared to other input devices and this is arguably a reason for their subsequent integration and popular-

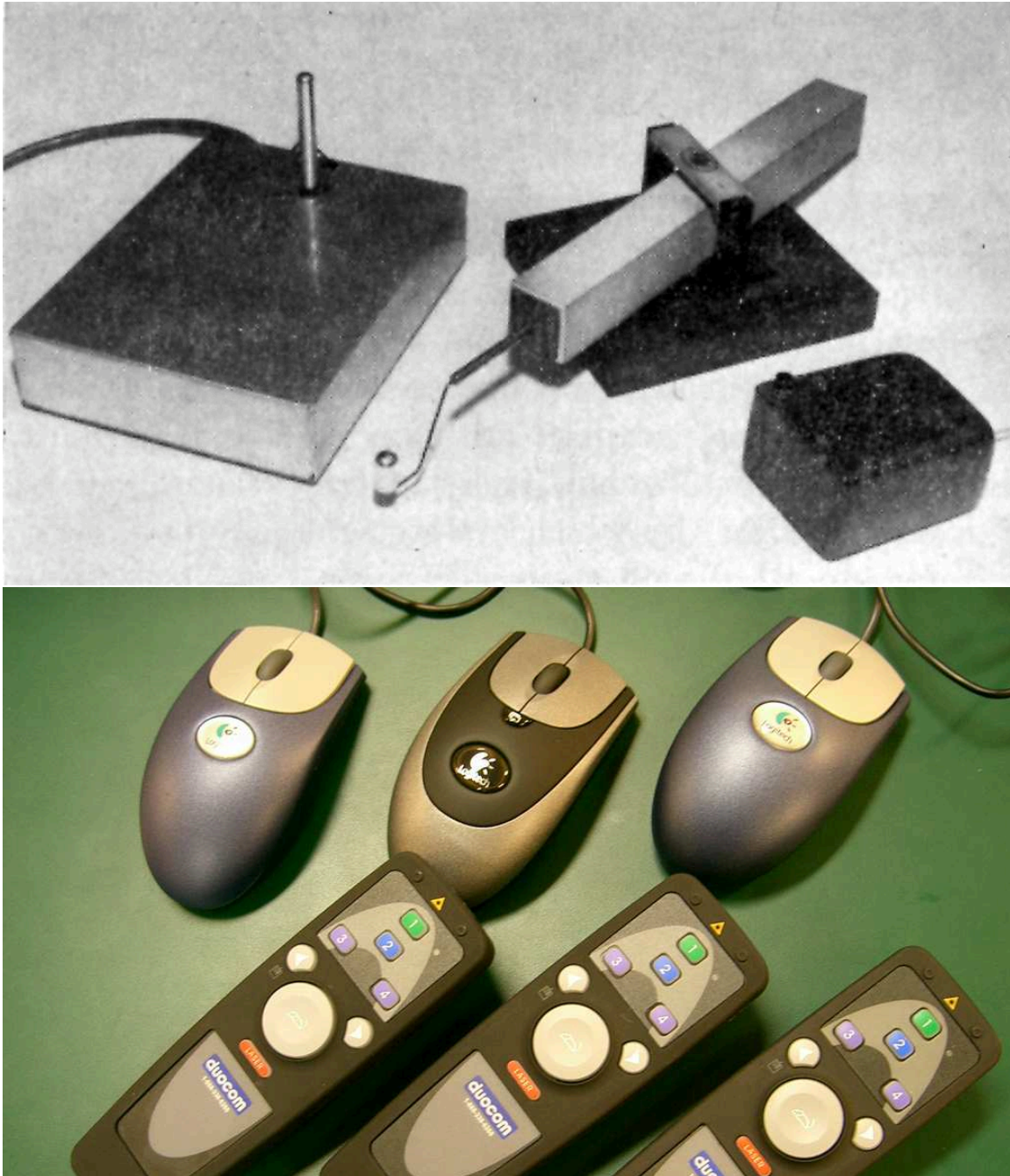


Figure 3.1: A collection of relative-mode input devices, including mice, joysticks, and trackballs. These input devices are characterized by the way in which they track user input. Unlike absolute-mode input, these devices use changes in motion to determine the location of user input. The top photograph is from English, Engelbart, and Berman (1967), showing a variety of early devices. From left to right: joystick, the Grafacon (an early device for curve tracing), and the mouse. The bottom photograph shows similar devices as they are sold on the market today. These kinds of devices are especially common in the world of desktop computing.

ity in early GUI-based systems. Mice have been studied for use in computer-aided design (CAD) systems, in comparison to other multimodal techniques such as voice and gaze, and have also been compared to other devices that have more degrees of freedom (Price, 1984; Hinckley, Tullio, Pausch, Proffitt, & Kassell, 1997; Hansen, Torning, Johansen, Itoh, & Aoki, 2004). Differences between types of users, such as adults and children, have been studied: Inkpen (2001) investigated children's use of drag-and-drop versus point-and-click interaction styles using a mouse to determine whether the choice of interaction style affected their performance in interactive learning environments.

Various refinements and compliments to mouse interaction have been proposed. These include attempts to increase the degrees of freedom available to the mouse, toolkits to support and extend mouse interaction, and variations on mouse movement characteristics. The Rockin'Mouse proposed by Balakrishnan, Baudel, Kurtenbach, and Fitzmaurice (1997) is a device with four degrees of freedom that has the same shape as a regular mouse with exception of a rounded bottom so that it can be tilted. The VideoMouse is a mouse that uses computer vision techniques to enable movement with six degrees of freedom (Hinckley, Sinclair, Hanson, Szeliski, & Conway, 1999). The cubic mouse uses a six-degree-of-freedom tracking sensor to specify 3D coordinate information inside interactive graphics applications (Frohlich & Plate, 2000). A toolkit to enable application support for interaction with multiple mice in collaborative settings has been implemented by Tse and Greenberg (2004) and an acceleration technique proposed by Baudisch, Cutrell, Hinckley, and Gruen (2004) addressed the issue of acquiring targets using a mouse in systems with multiple displays.

### 3.2.2 Pointers, Wands, and 3D Interaction

Though the mouse and WIMP-style metaphors are important in the desktop world, another class of input devices has evolved to support interaction with large screens and immersive environments. These devices are variously known as *pointers* or *wands* and many (but not all) support spatial interaction with at least three dimensions of positional movement. Many of these devices also track movement about three dimensions of orientation. For economy, we will refer to input devices with three or more degrees of freedom as devices that support *3D interaction*. One of the major implications of 3D interaction is the capacity to move beyond relative-mode input devices and toward *absolute-mode* input devices, or devices for which there is a direct correspondence between where one is pointing and the tracked

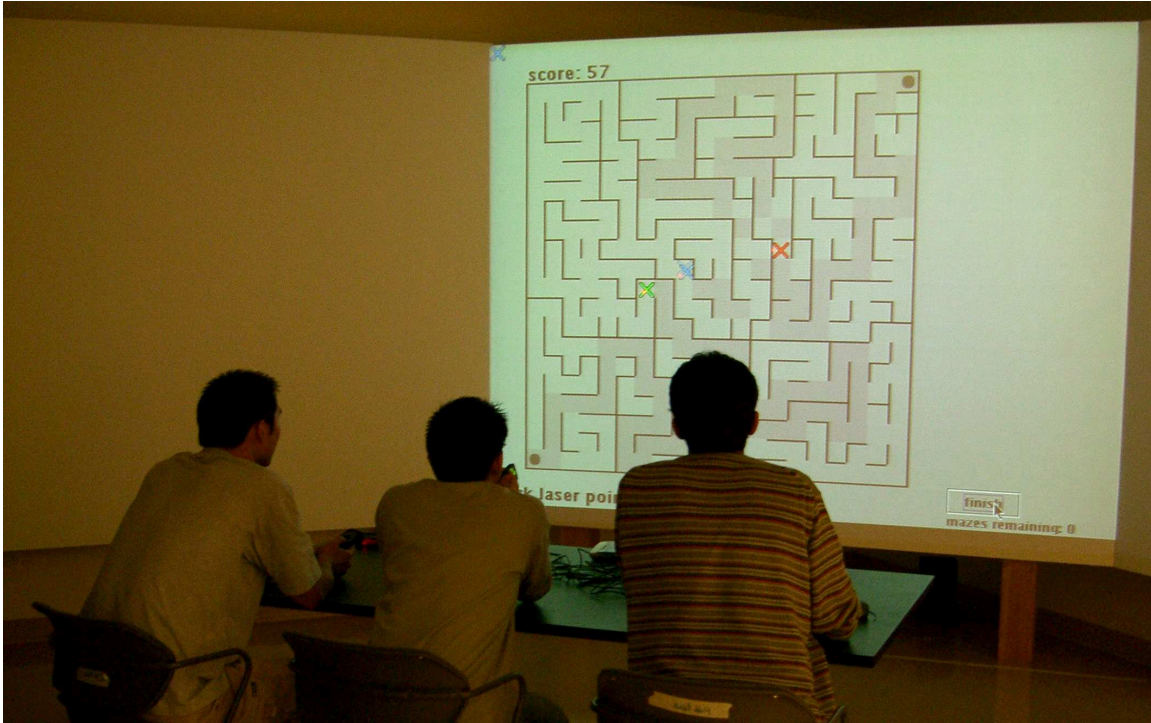


Figure 3.2: An example of laser pointer interaction with large screens, from Vogt et al. (2004). Laser pointer interaction is an example of pointers or wands in action. Such input devices can be used to support user interaction in scenarios where mice may be infeasible or undesirable. They also demonstrate the difference between relative- and absolute-mode input: the latter is characterized by a mapping between where a user points and the tracked position on the display.

position on a graphical display.

Figure 3.2 illustrates large screen interaction with laser pointers, which is one form of pointer or wand input. Laser pointers typically use computer vision and image processing techniques to identify the locations of projected points on the screen. This is a technique that has been used by Kirstein and Mueller (1998), Olsen and Nielsen (2001), Oh and Stuerzlinger (2002), Myers et al. (2002), and Vogt et al. (2004) among others. Vision-based tracking has the advantage of being inexpensive and scalable, although performance can suffer when the tracking is done under lighting conditions that are less than optimal. In situations more closely resembling immersive virtual environments, a multidimensional tracking system may be used, such as the electromagnetic Polhemus Fastrak (developed by Polhemus Incorporated) or the ultrasonic InterSense IS-900 tracker (developed by InterSense Incorporated). These technologies are generally more reliable, especially ultrasonic tracking, although these have the chief disadvantages of being fairly expensive and of being tethered

in the environment.

There are many issues surrounding the use of pointers and wands for 3D interaction. Hinckley, Pausch, Goble, and Kassell (1994b) review many of these, including perceptual issues, ergonomic constraints, interaction metaphors, and calibration. Ware and Osborne (1990) have suggested that there are at least three different kinds of 3D interaction: (1) the eyeball-in-hand, where the view the user sees is controlled by hand-guided manipulation of a virtual camera, (2) the scene-in-hand, where the user has an external (exocentric) view of an object and manipulates the object directly via hand motion, and (3) flying, where the user navigates a vehicle to move about the scene. Hinckley et al. (1994b) add a fourth kind: (4) ray-casting, where the user indicates the position of a target by casting a ray or cone into the 3D scene. This last kind of 3D interaction most closely resembles the kind of pointing done in WIMP-style GUIs, although it is clear that 3D interaction extends well-beyond what would be considered “typical” on a desktop.

There have been many kinds of pointer and wand extensions to ray-casting and direct pointing. Liang and Green (1993) implemented ray-casting by allowing the user to hold a pointer or wand in a comfortable position and rotating it to change the casted direction. A variation of ray-casting, known as “cone-casting” was also implemented by Liang and Green to permit the selection of multiple objects in a well-defined volume. Hinckley, Pausch, Goble, and Kassell (1994a) used a semi-transparent plane to select cross-sections of a polygonal brain model by extending the ray-casting metaphor into a “plane-casting” metaphor. Similarly, Bier et al. (1993) implemented a semi-transparent tool sheet for their Toolglass interface. Poupyrev, Billinghurst, Weghorst, and Ichikawa (1996) have further proposed the Go-Go interaction technique as a means for allowing users to acquire objects in a 3D environment that lie outside of their reach.

Numerous studies have compared the performance of pointers and wands to mice and similar indirect input devices. Among others, Poupyrev, Billinghurst, Weghorst, and Ichikawa (1998) have compared the performance of virtual hand interaction with that of virtual ray-casting and with the Go-Go interaction technique while Myers et al. (2002) compared the performance of laser pointers with several other input devices, including mice and a hybrid “semantic snarfing” technique. These and other studies have shown that pointers and wands are not Fitts’s Law-optimal input devices for target acquisition. Because these devices are not typically used on a surface like a tabletop, they are subject to disadvantages such as

---

hand jitter and the “pen-drop” phenomenon: button presses to indicate selection are typically in a direction orthogonal to that of pointer or wand movement, leading to systematic pointing inaccuracies at the moment of selection. Nevertheless, the value of pointers and wands may not come from empirical performance but rather from their flexibility. Pointers and wands can be valuable for collaboration, where gestural communication may be more important, and they can also be employed in situations where mice and other similar input devices cannot be used, such as in classroom environments (Vogt et al., 2004).

### 3.2.3 Pens, Styli, and Touchscreens

Touch-sensitive displays form another class of absolute-mode pointing devices. Unlike pointers and wands, these displays require direct contact for interaction. Touch input has become more popular as small, portable displays have become more important. “Touch” in this sense not only refers to direct contact with the finger, but also the use of implements such as pens and styli to make contact with a display. Because small devices make traditional desktop input cumbersome and 3D interaction techniques impractical at best, touch has become a particularly intuitive kind of interaction. Most users have ample experience pushing mechanical buttons and many of the graphical widgets found in modern GUIs lend themselves as much to direct touch as they do to indirect input with mice and similar devices.

Figure 3.3 shows an example of touchscreen technology in action. Although touch technology has generally received less attention in HCI compared to the mouse, it is possible to trace the history of touch input before that of the mouse. Ivan Sutherland, who is generally credited with envisioning the idea of virtual reality, worked on an interactive computer application called Sketchpad in the early 1960s (Sutherland, 1963). Sketchpad is famous for its use of the light pen, which allowed users to point at and interact with graphical objects on a screen. Unlike most other computer systems of the time, Sketchpad and its light pen were unique in allowing the user to directly interact with the system. The light pen had a profound influence on the role of the user in computing environments, and many of the terms that are now common in HCI, such as “icon” and “direct manipulation” are consequences of its impact (D. Smith, 1977).

In the context of pointing and target acquisition, there have been many studies of the effectiveness and usability of touch-based interfaces. Albert (1982) found that finger-based touchscreens were superior in speed and worst in accuracy when compared to a



Figure 3.3: An illustration of touchscreen technology at work. Touchscreens and other touch-sensitive displays have become more popular with the increased demand for portable devices, such as mobile phones and PDAs. Touch input primarily differentiates itself from other forms of absolute-mode input by requiring the user to make contact with the display surface (either directly with the finger, or with a physical implement such as a pen or stylus) to indicate interaction.

number of other graphical input devices. Potter, Shneiderman, and Weldon (1988) found that user performance with touch displays could be manipulated based on how the touch-based interaction was initiated. Sears and Shneiderman (1991) subsequently found that accuracy problems in earlier touchscreen studies may not have been due to users but rather to technological limitations because their study found that participants could accurately hit graphical targets as small as four pixels on a high-resolution display.

Empirical work has also looked at pen and stylus-based input, especially for handheld and tablet-size displays. Stylus-operated touchscreens were found to be comparable to a mouse when measured in terms of speed and accuracy in a study by Mack and Lang (1989). A survey of pen-based interaction by Citrin et al. (1993) focused on the potentials and limitations of such input for personal workstations, characterizing the bandwidth of pen input as being at least comparable to that of mouse input. Two experiments by Ren and



---

Moriya (2000) empirically evaluated item selection tasks using pen-based input, suggesting that such input could be modeled as state transitions. Another study by Hancock and Booth (2004) looked at pen-based menu selection across horizontally-oriented displays, such as tabletops, and more conventional vertically-oriented displays suggesting that user performance with pen input not only varied with display orientation but also with handedness of the individual. Other work by Parker, Mandryk, and Inkpen (2004) has demonstrated that it is possible to integrate distal pointing (such as laser pointer interaction) with touch pointing and that this can be valuable on tabletop displays under conditions where objects for selection may initially be out of reach.

### 3.2.4 Voice and Multimodal Input

Multimodal interaction, or interaction by combining more than one kind of input or output at the same time, is becoming increasingly important as senses other than vision and actions other than pointing are exploited in the user interface. Although it is possible to argue that such interaction has been an important goal in HCI for many years, multimodal input must contend with technical, theoretical, and perceptual issues that have yet to be fully explored or understood (Sutherland, 1965; Bolt, 1980). Nevertheless, there are many who believe that interfaces supporting multimodal interaction are inevitable and that such interfaces represent an evolutionary “next step” in user interface design (Oviatt & Cohen, 2000).

Notebook computers with handwriting capabilities (i.e. PDAs and Tablet PCs) are now commonplace and many of these include some kind of voice recognition and speech production software for textual transcription and playback. Many perceptual user interfaces (PUIs) are tangible examples of interfaces supporting multimodal input, and physiological measurement devices such as eye movement monitors and skin response sensors are now being used to infer the internal attentive and affective states of the user (Picard, 1997). Other work into tangible user interfaces (TUIs) seeks to blur the distinction between the physical world and the graphical world and can be seen as multimodal interaction because of its tight integration of haptic cues with more traditional visual cues (Ullmer & Ishii, 2002; Jacob, Ishii, Pangaro, & Patten, 2002).

As part of the multimodal theme, voice input has been an input method of considerable interest, although few practical systems exist. Even though there are claims that voice recognition technology has achieved accuracy ratings of 95 percent or better, these are still

insufficient for true, fluid interaction since 95 percent accuracy corresponds to one mistake in every twenty words, which is still unacceptable for many potential applications (Oviatt & Cohen, 2000). Other empirical studies of voice input compared to other kinds of input, such as using a mouse, have not shown voice to be any faster than conventional motor input (Christian, Kules, Shneiderman, & Youssef, 2000). Nevertheless, providing voice as an option permits flexibility under certain circumstances, especially with respect to issues of universal accessibility. The non-verbal aspects of vocal input, such as the attributes of pitch and volume, have been shown to have some value as inputs for interactive applications (Igarashi & Hughes, 2001).

### 3.3 Graphical Displays and Visual Output

The visual display comprises the output component of an interactive system and makes up much of the visual input to the perceptual system of the user. It is fair to say that graphical displays have evolved alongside that of input devices and technological advances on the output side have necessitated similar advances on the input side. Many of the major milestones that are linked to the development of the graphical display can also be linked concurrently with the invention of specific input devices. Sutherland's (1963, 1965) Sketchpad light pen was first used on a 2D vector display, which was among the first true graphical computer displays. Engelbart's (1963) invention of the mouse is linked to the development of computer displays for supporting real-time text editing. As WIMP-style GUI systems such as the Xerox Alto and Apple Lisa were being developed, the quality of the graphical display was one of prime importance. The preliminary design of the Xerox Alto in 1972 included a "bit-mapped display screen measuring 606 pixels horizontally by 808 pixels vertically, thus producing an upright rectangle about the size and shape of a standard  $8\frac{1}{2}$ -by-11-inch sheet of paper" (Waldrop, 2001). Even many years later, displays with an effective 800 by 600 pixel resolution are standard for many graphical applications.

Several taxonomies exist with respect to categorizing graphical displays although it is arguable that these categorizations have received limited attention relative to those in the realm of input devices. The ACM SIGCHI Curricula for HCI (1992) defines output devices according to their mechanical attributes, performance characteristics, and accessibility. Milgram, Takemura, Utsumi, and Kishino (1994) have proposed a taxonomy of

---

visual displays for augmented reality (AR) systems. Nesbitt (2004) has further proposed the MS-Taxonomy, which is a classification of abstract data displays that is general for all sensory modalities. In the discussion that follows, we categorize visual displays according to their size and form factor. This presents graphical displays from the perspective that displays of different scales lend themselves to different kinds of activities and also have unique technical and perceptual challenges.

### 3.3.1 Monitors and Desktop Displays

Monitor-scale displays are the most common kinds of graphical displays and owes their popularity to the success of the desktop computer and early time-sharing systems. Desktop monitors vary little in size, typically ranging from 15 inches to 21 inches diagonally. At present, the two most common technologies for desktop monitors are cathode ray tube (CRT) displays and newer liquid crystal displays (LCD). CRTs generally have better contrast when compared to equivalent LCDs but take up more space and can be prone to geometric distortions. Regardless of display type, current desktop monitors are capable of achieving resolutions of 1600 by 1200 pixels or greater.

Research into desktop display technology in computer graphics and HCI has focused on improving the image quality and the ergonomics of these displays. Fish-tank virtual reality (VR) refers to the use of a desktop monitor along with measurements of head and eye position to provide a correct perspective view of a small virtual environment (Ware, Arthur, & Booth, 1993). Fish-tank VR displays are frequently coupled with stereo imaging technology to produce binocular imagery and have been the focus of several studies looking at the overall effectiveness of stereoscopic graphics and dynamic head-coupled perspective projections, also known as head-tracking, for 3D navigation and wayfinding (Arthur, Booth, & Ware, 1993; Ware & Rose, 1999).

Other work has looked at improving the colour range of desktop-scale displays. Many current generation graphics cards now natively support floating-point precision colour discrimination, which has led to the development of high-dynamic range (HDR) displays (Seetzen et al., 2004). Such displays provide greater contrast and colour expression than standard desktop monitors, CRT or otherwise, and have applications in entertainment and information visualization. A prototype desktop-scale display prototype with multiple focal planes is also a promising direction for desktop displays (Akeley, Watt, Girshick, & Banks, 2004).

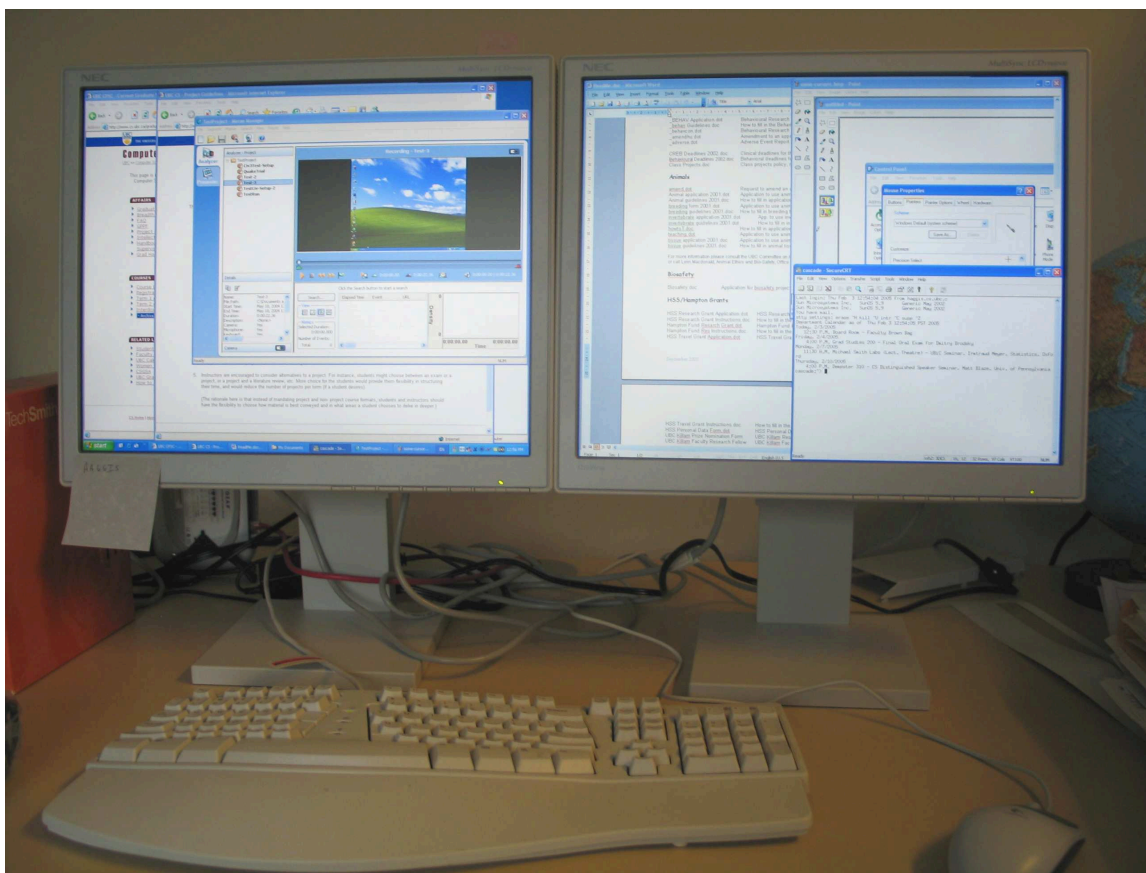


Figure 3.4: A desktop system with multiple monitors. Many GUI systems now have support for multiple displays, which has become a topic of some interest in HCI. Multiple monitors allow flexibility and extensibility in the organization of WIMP-style user interfaces although the perceptual impact of multiple displays remains an open question.

One can imagine extending existing fish-tank VR technology with HDR stereoscopic graphics implementing correct focal cues, which remains one of the major technical limitations of current stereoscopic techniques.

Figure 3.4 illustrates a desktop environment with a multiple monitor configuration. Augmenting desktop computing with more than one graphical display is another area of research important to HCI. It remains an open research question how the perceptual impact of multiple displays should be characterized. Grudin (2001) suggests that at least one dimension of this lies in the differentiation of foveal and peripheral perceptual cues offered by multiple monitor arrangements. Ringel (2003) offers a taxonomy of organization strategies based on her observations of how users tend to spatially arrange multiple desktop workspaces. These are supported by findings from Hutchings and Stasko (2004), who point

---

out that the value of multiple monitors does not necessarily lie in supporting interaction with additional information but rather in helping users maintain situational awareness.

### 3.3.2 Large Screens and Immersive Displays

Large screens are a class of graphical displays that are of particular interest to those studying VR and collaborative computing environments. Unlike desktop displays, large screens can often be measured in feet rather than inches and are usually fixed installations that cannot be easily moved. Such displays may require separate projection units, which could be CRT-, LCD-, or Digital Light Processing (DLP)-based. Because of space constraints, some large screen facilities may have these projectors forward-mounted (i.e. in front of the display surface), which saves space but runs the risk of shadow-casting and occlusion effects by users standing too close to the display. Rear-mounted (i.e. behind the display surface) projectors do not suffer from these effects but usually require expensive mirrors on a substantially larger projection area behind the screen. Unlike desktop displays, large screens may require frequent maintenance, especially with respect to the calibration and registration of the displays (Raskar et al., 1999). This is a situation that is common when large displays are tiled together with multiple projectors.

The CAVE and similar multi-screen display environments as illustrated in Figure 3.5 are excellent examples of how large screen technology is often used (Cruz-Neira, Sandin, DeFanti, Kenyon, & Hart, 1992). Although many large screen configurations consist of a single screen practical for meeting rooms and many classroom environments, multiple screens can be physically arranged to provide what is frequently called an immersive virtual environment. These environments are used for a variety of tasks, including 3D navigation and wayfinding, large-scale visualization, and vehicle design among others (Kasik, Troy, Amorosi, Murray, & Swamy, 2002; Swindells et al., 2004). Like their fish-tank VR counterparts, real-time measurements of head and eye position are used to calculate proper perspective projections and stereoscopic imagery for a single user.

Large screen interaction has become important outside of VR in the context of computer-supported cooperative work environments where it is used for Single Display Groupware (SDG) applications (Stewart, Bederson, & Druin, 1999). Several research systems use large screen displays to share information among multiple users (Cavens, Vogt, Fels, & Meitner, 2002; Izadi, Brignull, Rodden, Rogers, & Underwood, 2003; Khan, Fitzmaurice, Almeida,



Figure 3.5: A user working inside a CAVE-style multi-screen display environment (Swindells et al., 2004). Large screens are typically used in this fashion as the basis for immersive virtual environments for 3D navigation and wayfinding. These displays are expensive and evidence for their utility remains an open question in HCI. Large screen display configurations differ upon context of use. Some consist only of a single screen while others still are horizontally-oriented in the form of tabletop displays.

Burtnyk, & Kurtenbach, 2004). Others have used large surfaces in a horizontally-oriented format as a means for enabling collaborative interaction around a tabletop (Dietz & Leigh, 2001; Shen, Vernier, Forlines, & Ringel, 2004; Ringel, Ryall, Shen, Forlines, & Vernier, 2004). In all of these cases, there are issues of how to protect the privacy of individual users as well as how designers should best provide the means for enabling all of the participants to communicate equally in a collaborative large screen scenario (Tan & Czerwinski, 2003; Vogt et al., 2004).

Perhaps one of the largest research questions surrounding large screen displays is measuring their usefulness. Because it can be rather difficult to define the general characteristics of a useful display, much more a large display, there is little quantitative evidence suggesting that large screens confer any substantial advantage performance-wise (Kasik et al., 2002; Swindells et al., 2004). One study by Pausch, Proffitt, and Williams (1997) has even sug-



Figure 3.6: A selection of small screens and handheld displays, including PDAs, mobile phones, and Tablet PCs. A chief issue surrounding their effectiveness is the limited amount of display real estate. Unlike desktops or large screens, graphical information like Web pages must be condensed into a representation that loses as little of the original intent as possible.

gested that exposure to an immersive display could create a “negative learning transfer effect” when moving from an immersive display to a smaller display. Nevertheless, a study by Czerwinski, Tan, and Robertson (2002) suggests that the peripheral context provided by a large display may facilitate improved navigation and wayfinding performance for women (but not for men) under certain conditions. Subsequent work by Czerwinski et al. (2003) has looked at other methods for characterizing the benefits of large displays, including measurement of cognitive load and peripheral effort spent on display organization instead of actual task completion.

### 3.3.3 Small Screens and Handheld Displays

Small and portable screens, such as those found on PDAs, mobile phones, and some notebook computers like Tablet PCs form another class of graphical displays (see Figure 3.6 for examples). These kinds of displays are unique because the devices for interacting with the display are usually built into the display itself (i.e. they may be touch-sensitive or have

a built-in keyboard) and because attributes other than display quality such as battery life and portability may be more important. These displays may range in size from one inch to twelve or more inches diagonally, depending on the device. LCDs are the most common technology for these displays, although new technologies such as organic light emitting diodes (OLEDs) are rapidly becoming viable alternatives.

Handheld devices are an integral component of the overall vision for ubiquitous computing, where a computing device of some kind is available to users everywhere and all of the time (Want et al., 1995). One topic of past research has been applications for small devices that may have access to a networked infrastructure. Pham, Schneider, and Goose (2000) have used a situated computing framework to support small device access to libraries of rich multimedia content not typically available on small screens. Buyukkokten, Garcia-Molina, Paepcke, and Winograd (2000) have designed and implemented a Web browsing facility for small screens that have limited capabilities. Kawash (2004) has proposed the use of “declarative” user interfaces as a means for automatically consolidating the representation of Web information as it is displayed on mobile devices with different form factors.

One of the major issues surrounding the use of small screens is how best to use the limited amount of screen space available to the visual interface. Many techniques have been proposed to efficiently view and manipulate information on these kinds of devices. The metaphor of stretching a rubber sheet has long been suggested as a metaphor for viewing large layouts on small screens (Sarkar, Snibbe, Tversky, & Reiss, 1993). The Rapid Serial Visual Presentation (RSVP) technique has been used to facilitate the process of browsing the Web (de Bruijn, Spence, & Chong, 2002). Unlike the rubber sheet metaphor, which organizes visual information spatially, RSVP rapidly presents individual frames of visual information to the user over time. An empirical evaluation of information visualization techniques used on larger displays, including fisheye, zoom, and panning, found that both fisheye and zooming techniques were viable on small displays but panning performance was less compelling (Gutwin & Fedak, 2004). A segmented zooming technique known as ZoneZoom has also been implemented for map navigation tasks on mobile phones, suggesting that such techniques can have a substantial impact on the effectiveness of even very small displays (Robbins, Cutrell, Sarin, & Horvitz, 2004).

Another research direction for small displays has involved integrating their use with larger displays, such as desktop- or even large-scale displays. Myers et al. (2002) have



proposed the technique of “semantic snarfing,” which involves gross selection of objects on a large screen and finer manipulation on an accompanying small device. Yee (2003) has suggested combining pen input with external spatially-aware displays to enable a kind of two-handed interaction. Hinckley (2003) has also explored the use of synchronous gestures as a means for dynamically tiling together small displays to create a single, larger display via a combination of touch and tilt sensors.



## Chapter 4

# A Representational Approach to HCI

In the previous two chapters, we discussed the concepts of mental representation and human vision as they relate to a computational theory of the human mind. We also reviewed some of the input and interaction technologies that might benefit from an understanding of mental representations. In the next three chapters, we will discuss the experimental approach that will yield the basis for understanding the integration, separation, and mediation of visual perception and motor action from a representational perspective. The common experimental methodology for the three representational theories of stimulus-response compatibility, the upper and lower visual fields, and the two-visual systems hypothesis discussed in this dissertation is the basis of a theoretical framework for a *representational approach* to evaluating user performance and is a means for researchers to extend and augment the empirical evaluation methods already in common use in the study of HCI.

### 4.1 Describing the Theoretical Framework

The psychological literature suggests that mental representations are a means for understanding the ways in which visual information is processed for the purpose of enabling motor movement. The HCI literature suggests that visually-guided motor movement has traditionally been, and will continue to be, an important part of user interaction. It is the *relationship* between visually-guided movement as it is studied in cognitive psychology and related fields such as kinesiology, and visually-guided movement as it is studied in HCI that is of interest in this dissertation. Although it is apparent that both researchers and practitioners alike in HCI value the study of vision and movement, the evaluative approach taken thus far has heavily emphasized the performance characteristics as described by the level of computational description, with little attention to the reasons as to *why* and *how* such char-

acteristics exist as described by the levels of representation and hardware implementation (Marr, 1982).

The representational approach taken in this dissertation is an extension of research practices already common in HCI. The major difference between this particular approach and other approaches lies in the body of theory that is considered appropriate for study. The approach can be outlined as a six-step process:

1. Identify a representational theory of human behaviour in the psychological literature that could explain how and why certain behavioural phenomena occur.
2. Concurrently identify existing user interface design and user performance issues in the HCI literature that are directly related to predictions of a particular representational theory of behaviour.
3. Based on previous work in psychology and HCI, generate hypotheses about how the representational theory can predict user performance.
4. Devise a controlled experiment that can be conducted in an appropriate context of use to test these hypotheses.
5. Collect and analyze experimental data from participants in a manner consistent with the experimental design.
6. Interpret the experimental results to explain user performance characteristics, learn about the elements of an effective user interface, and derive practical design guidelines.

This approach closely resembles the way in which many experimental studies arise in HCI. The six-step process itself is very similar to one part of the “reflective practitioner” model described by Schön (1983), which emphasizes the importance of theory-driven design and how theory can effectively transform practice. In the case of this dissertation, the representational theories provide a new perspective on user performance that can be used to influence interface design practices. It is not the six-step process but the representational theories being examined that are especially novel.

---

## 4.2 Representations for Perception and Action

This dissertation demonstrates that *a representational approach to user performance is able to predict variations in visually-based user performance neglected by existing theoretical frameworks in HCI and that such insight can be helpful in the design and evaluation of interactive systems.* The representational approach described above can be applied to any number of different representational theories and is not especially limited to those that deal with human vision. To demonstrate the utility of mental representations as a theoretical framework for HCI, this dissertation has chosen three particular representational theories that have a direct connection to visually-based user interaction as evidence that a representational approach to HCI has practical merit. These theories are all connected in the sense that they demonstrate how the same visual information perceived by a user can be simultaneously interpreted in three different ways to yield different user performance characteristics that are important for different visual aspects of a user interface.

In cognitive psychology, it is postulated that the transformation of visual information into an appropriate motor movement is mediated by multiple mental representations of visual space. Figure 4.1 presents a representational model of perception and action as it is currently understood in cognitive psychology, derived from an illustration from Prinz and Hommel (2002). Their model summarizes anatomical and experimental evidence that different kinds of cognitive-sensorimotor interactions may arise from different representational modules mediating between perception and action. This model attempts to unify anatomy and observed phenomena in a way that resembles Marr's (1982) representational framework and the model serves as a convenient unifying framework for discussing the major representational modules in this dissertation.

In Figure 4.1, the three major representational modules examined in this dissertation are noted in italics and are illustrated in relation to the particular functions they serve in the model. Stimulus-response (SR) compatibility, the upper and lower visual fields (UVF/LVF), and the two-visual systems hypothesis can be seen as representational modules of perception and action that integrate, separate, and mediate visual information perceived by a user and result in actions performed by the user.

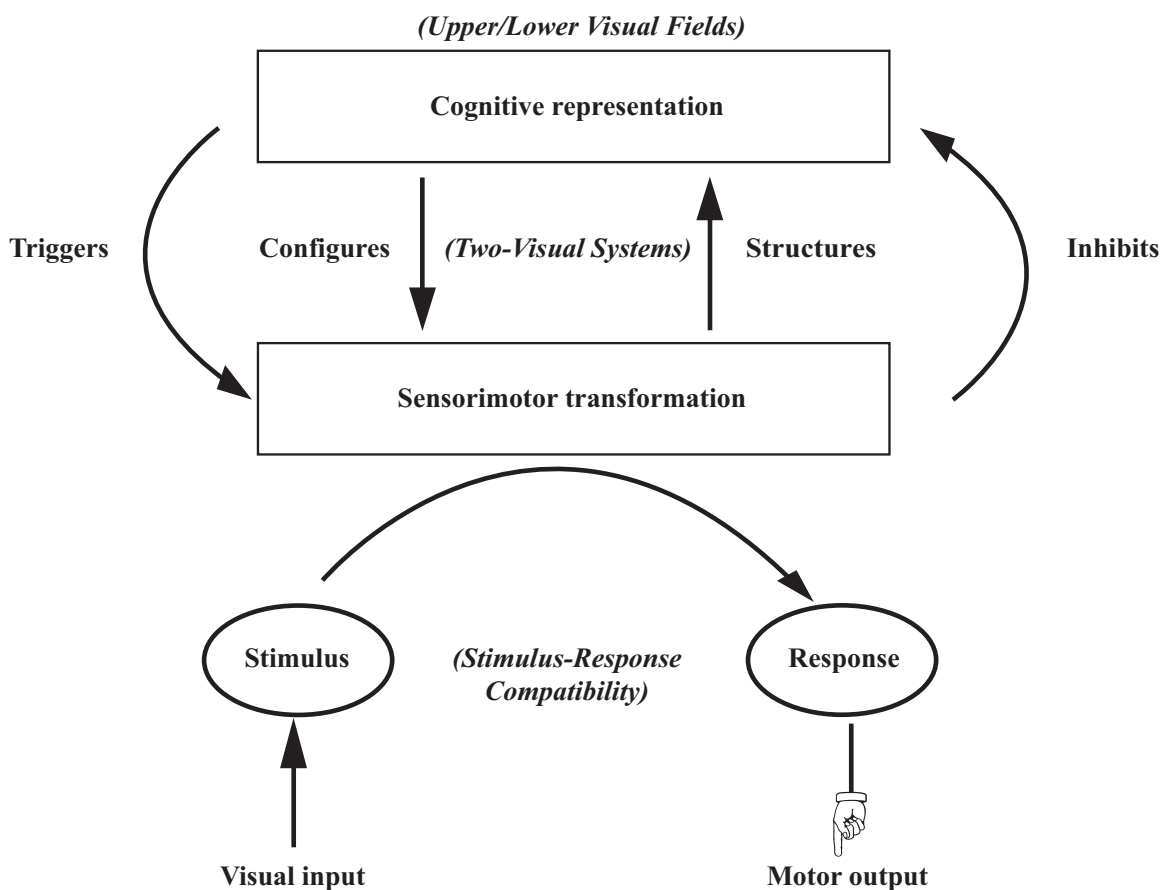


Figure 4.1: A summary of the interactions observed between the sensorimotor and cognitive representations of a stimulus leading to a behavioural response, based on an illustration from Prinz and Hommel (2002). The three representational modules explored in this dissertation (noted in italics) are described in relation to Prinz and Hommel’s particular model of perception and action.

#### 4.2.1 Integration of Visual Information

One way to look at how visual information is used for movement is by looking at vision for movement from an *integrated* perspective. The integration of visual information in this sense refers to the mapping between perceptual and action attributes and the effects this mapping has on response performance. The bottom of Figure 4.1 suggests that visual stimuli map directly onto motor outputs in a manner that affects the way in which visual information is transformed for movement. There is a body of literature in cognitive psychology and kinesiology that postulates a specific representational model that integrates visual information in this way. This is known as the study of stimulus-response (SR) compatibility, which we examine in Chapter 5.

### 4.2.2 Separation of Visual Information

Another way to look at vision for movement is from a *separable* perspective. The separation of visual information refers to representational theories that define perception and action as separate attributes of vision. Thus, in these kinds of representational theories, perception and action are independent mental representations themselves and visual tasks can be classified by whether they are perceptual in nature, such as object identification, or whether they are sensorimotor in nature, such as pointing. The top of Figure 4.1 suggests that different representations can trigger, inhibit, and configure appropriate motor responses. Both anatomical and experimental evidence exist to suggest that certain representations can be “activated” by factors such as perceived spatial location. One theory that does this is the functional specialization of the vertical visual hemifields, often known as the upper and lower visual fields (UVF/LVF), which we examine in Chapter 6.

### 4.2.3 Mediation of Visual Information

The middle of Figure 4.1 indicates that there is a bidirectional relationship between representations of visual space and the transformation process from stimulus to response. The illustration suggests that mental representations configure motor movements and that movements also structure the representation. Thus, another way to look at how visual information is used for movement is by looking at vision for movement as a *mediated* process. One question that can be posed about mental representations from this perspective refers to what mechanisms human vision uses to decide when a visual task is inherently perceptual or active in nature. Over thirty years of evidence in experimental psychology has led to the development of what is known as the two-visual systems hypothesis, which is a representational theory of perception and action describing mechanisms that predict visual behaviours consistent with the mediation of visual attributes for movement. We examine this representational module in Chapter 7.

## 4.3 Implications for HCI

When we think about the ways in which users interact with computer systems, we arrive at the inescapable conclusion that physical movement, whether it be in the form of pointing with a mouse in a WIMP-style GUI or physical interaction in a more immersive setting,

will always be an important part of HCI. The term “immersive” is used as a general term to describe computing environments that are situated in a wider physical context, much as it is meant when talking about ubiquitous or pervasive computing environments. The visual attributes of perception and action are critical, inescapable elements of user interaction that have thus far been implicit in understanding the characteristics of user performance. In the limited realm of desktop computing, a deeper understanding might not seem important (although as we will see, representational theories can still yield valuable insight into desktop interaction). However, with HCI moving beyond the desktop, understanding the lower-level mechanisms that drive user performance may be unavoidable. The paradigm of direct manipulation, which has mostly been applied as an interaction metaphor for the indirect control of menus or cursors displayed on a screen by mice or other devices held in the hand, is giving rise to truly direct interaction in the form of technologies that give natural, real-world meaning to physical movements like distal pointing, gesture, reaching, and grasping.

The representational framework introduced here is a means to complement our existing understanding of user performance. It is perhaps fortuitous that the experimental studies of cognitive psychology and kinesiology have emphasized the use of elementary movement tasks like pointing, reaching, and grasping, since these also form the basis for user interaction. Experimental work into visually-guided movement in cognitive psychology suggest that systematic variations in movement performance exist, which can only be accounted for by a representational account of visual processing. The three representational theories discussed in this dissertation will demonstrate three such instances of systematic variation in the context of HCI. Because existing evaluation techniques in HCI have generally limited themselves to a purely descriptive account of performance, such performance variations have not been central to research in HCI. Some may attribute such variations when they are detected as unexpected phenomena and may account for them in ways already predicted by representational theories. In the worst case, performance variations may be dismissed as statistical or experimental error when in fact they may be symptoms of something intrinsically deeper about user performance.

The study of HCI has traditionally required the application of a wide variety of evaluative methods, ranging from empirical studies in ecologically-valid scenarios to controlled experimentation in lab settings. A theoretical framework driven by theories of mental representation is readily applicable to all methods of study in the spectrum of evaluative



techniques. For example, theories of knowledge representation could be used in qualitative case studies to understand user patterns in collaboration while theories of visual representation could be used in controlled experiments to understand lower-level characteristics of user performance. For the purposes of maintaining a unified theme throughout this dissertation, we focus on the latter by applying the three representational theories to investigate user performance phenomena through controlled experimentation.

A representational approach may also allow researchers to more rigorously explore the ideas of mental models and affordances, which are commonly applied in HCI. In the study of HCI, there is an apparent disconnect between “theoretical” empirical measures of user performance and the means for facilitating an enjoyable, effective user experience by connecting interaction with intuitive experience. Unification of the observable experience with those phenomena that are not apparent or consciously perceivable is central to the computational theory of mind and mental representations. As HCI continues to study the characteristics that constitute “natural” user interaction, it can be useful to study how users understand interaction in the natural world; this is conveniently captured in the theoretical framework of mental representations. A representational approach to HCI could eventually lead to a more precise understanding of design elements that make good interactive systems enjoyable and effective.



## Chapter 5

# Stimulus-Response Compatibility<sup>1</sup>

The integration of visual information for perception and action is observed in the behavioural phenomenon known as stimulus-response (S-R) compatibility. Because of its explanatory and predictive power, it is a phenomenon of some interest in several disciplines, including cognitive psychology, ergonomics, and HCI. S-R compatibility encompasses an entire class of observed behaviour, although it is most often described as a spatial phenomenon involving the congruency between the position and orientation of a presented visual stimulus and its associated response. A classic experiment by Brebner (1979) includes the presentation of two lights (a left light and a right light) with two corresponding keys for response (a left key and a right key).

Those S-R combinations that are seen to facilitate movement performance are typically referred to as *compatible* mappings (i.e. using the left key to respond to the left-sided light and using the right key to respond to the right-sided light) while those that are seen to interfere with performance are typically referred to as *incompatible* mappings (i.e. using the left key to respond to the right-sided light). S-R compatibility may also be characterized as a symbolic phenomenon, since its effects extend to other sensory modalities like audition (i.e. a deepening pitch is observed to be compatible with downward movement while a rising pitch is observed to be compatible with upward movement), or as a dynamic phenomenon, since temporal S-R effects are also known to exist (i.e. presenting an example of a complex movement is known to improve subsequent performance of that movement) (Biel & Carswell, 1993; Sturmer, Aschersleben, & Prinz, 2000).

---

<sup>1</sup>A version of this chapter has been published. Po, B. A., Fisher, B. D., and Booth, K. S. (2005). Comparing cursor orientations for mouse, pointer, and pen interaction. Proceedings of the ACM Conference on Human Factors in Computing Systems, pp. 291-300.

## 5.1 Background and Related Work

In cognitive psychology, understanding the common coding process of visual information for perception and action in a shared medium (i.e. mental representation) is one of the key goals for studying S-R compatibility. Some of the earliest experimental work into S-R compatibility was conducted by Fitts and Seeger (1953), who were interested in describing systematic manipulations of motor performance using the mathematical terms of information theory. They are generally credited with originally defining the notion of compatibility between S-R relations. Subsequent experimental work by other cognitive psychologists such as Morrin and Forrin (1962), Simon and Rudell (1967), and Broadbent (1971) led to what may be informally known as the basis for the *coding hypothesis* of S-R compatibility. Brebner (1979) writes that “it is the complexity of the recoding process by which a particular stimulus is mapped onto its required response which determines the slowness, inaccuracy, and lack of processing capacity which typifies performance with incompatible relations.”

Since then, many different extensions to the original coding hypothesis have been postulated. Proctor and Reeve (1990) provide a substantial review (some of these are reviewed here). Bauer and Miller (1982) found that compatibility effects were dependent on the horizontal or vertical orientation of visual stimuli, suggesting that direction of movement was an attribute of the coding process. Experimental evidence from several other researchers found that the automatic priming of congruent S-R attributes could be explained as the result of dimensional overlap, or the degree to which S-R attributes have physical or semantic features in common (Kornblum, Hasbroucq, & Osman, 1980; De Jong, Liang, & Lauber, 1994). These results were later formalized into a referential-coding explanation by Lippa (1996), who found evidence supporting the claim that motor responses are mentally represented in reference to hand posture so that physically orthogonal S-R dimensions can overlap. Other evidence suggests that an alternative salient-features coding principle may be more appropriate: the translation between stimulus and response is faster when decidedly salient S-R features correspond (Proctor & Reeve, 1990; Proctor, Dutta, Kelly, & Weeks, 1994). Hommel (1995) has postulated that many observed S-R compatibility effects may further be the result of a dual coding process conceptualized as either a serial search through S-R representations or as the outcome of a conflict between response codes activated in parallel.

---

A particularly fruitful thread of research related to S-R compatibility involves the compatibility of combining spatial and non-spatial attributes. This is typically known as the Simon Effect, which shows that the spatial characteristics of a task can inadvertently affect the performance of a decidedly non-spatial task (Simon, 1969). Original experimentation by Simon (1969) and Craft and Simon (1970) found that response times toward a stimulus were faster when compared to movement away from a stimulus, even though the correctness of a response may have had nothing to do with spatial movement (i.e. indication of a particular colour or responding to a particular auditory stimulus). Because of the Simon Effect, Smith and Brebner (1983) found evidence suggesting that S-R compatibility may necessarily incorporate multiple recoding processes as final determinants of response time performance. Stoffer (1991) used the Simon Effect to suggest that attentional focus is a prime factor in determining S-R compatibility, but other work by Weeks, Chua, and Hamblin (1996) was unable to replicate his results, calling into question the validity of this claim. Kerzel, Hommel, and Bekkering (2001) review and report on experimental work involving the relationship between the Simon Effect and temporal presentation, otherwise known as stimulus onset asynchrony (SOA). Their findings are interpreted as direct evidence that visual information for perception and corresponding information for action are linked together in a single representational mapping.

Other psychological research has been influenced by the study of S-R compatibility. Although S-R compatibility is typically studied as a cognitive phenomenon, Michaels (1988) has interpreted it as an ecological phenomenon, citing evidence suggesting that the congruency of S-R relationships may be a property of implicit affordances between response and intended destination. Biel and Carswell (1993) examined S-R compatibility and the relationship between musical notations and (piano) keyboard performance, suggesting that non-traditional horizontal notations might be used to support novice performances with a keyboard instrument. Furthermore, Chua (1995) has used a dynamical systems approach to extend the qualitative characteristics of S-R compatibility to understand the interaction between perception and action in terms of movement coordination as a function of physical laws and principles.

### 5.1.1 Previous Work in Ergonomics

The study of ergonomics and human factors has traditionally been an application area for S-R compatibility. The facilitation of movement performance has always been an element of particular interest in industrial settings where efficiency of operator performance is important and time or safety may be critical factors (Chapanis & Lindenbaum, 1959). As such, there is a body of related ergonomics literature that has looked at S-R compatibility from a more performance-oriented perspective. Perhaps one of the more important reasons that S-R compatibility has received substantial attention in ergonomics relates to the robustness of the behavioural phenomenon. The psychological literature consistently reports that S-R compatibility effects cannot be trained away, suggesting not only that stimulus position cannot be excluded from action control, but also that S-R compatibility must be considered in the design of any efficient user-operated machine (Dutta & Proctor, 1992).

In ergonomics, S-R compatibility effects are often studied in the context of their influence on directional performance and the relative spatial organization of machine interfaces. Worringham and Beringer (1989, 1998) extend the dimensional overlap theories of Kornblum et al. (1980) and the salient-features model of Proctor and Weeks (1990) to suggest that at least three different design principles might be taken into account to preserve directional S-R compatibility: (1) control-display compatibility, or the relationship between the location and orientation of an operator relative to the control mechanisms for a given piece of equipment, (2) visual-field compatibility, or the preservation of aligned motion between the relevant limb segment and controlling element, and (3) muscle-synergy compatibility, or the employment of muscle activation patterns like flexion or extension and appropriately associating these with the visually-specified direction of movement. In a series of experiments, they found that visual field compatibility provided substantially more robust performance effects relative to the other two compatibility principles.

Other work by Chua, Weeks, Ricker, and Poon (2001) suggests that the spatial compatibility of the relative orientation of an operator might be alternatively described between levels of mapping, configuration, and global relations. Their experimental work concurs and extends the results of Worringham and Beringer (1998), by indicating that it may be possible to identify compatibility mappings in the design of a system not only in terms of direct, physical relationships between operator and machine, but also in terms of inter-

---

actions between multiple levels of spatial organization akin to how widespread these S-R compatibility effects might be.

Despite the presence of “theoretical” design guidelines for maintaining S-R compatibility, it has been suggested that applying these guidelines in practice is especially difficult. Payne (1995) found that the judgments of compatibility in naive participants between different S-R mappings were not very accurate, suggesting that designers might not be able to predict whether one particular control-display configuration might lead to better performance than another. Vu and Proctor (2003) subsequently found that such judgments could be improved with practice, but that common intuition alone was insufficient to make good design decisions. Tlauka (2004) further suggests that appropriate judgment of compatible configurations is a function of design difficulty. In his experiments, he found that simple compatible mappings were chosen with a high degree of accuracy, but that scenarios not solely relying on spatial congruence for optimal compatibility were substantially less accurate.

### 5.1.2 Previous Work in HCI

Some early work in HCI has benefited from the study of S-R compatibility, although it has received far less attention in recent years. John, Rosenbloom and Newell (1985) applied a computational description of S-R compatibility to formulate a GOMS model of user performance as it relates to presenting and remembering computer command abbreviations. Subsequent work by John and Newell (1989) extended the Model Human Processor of Card, Moran, and Newell (1983) to the domain of transcription typing. Both of these models were found to be highly reliable, serving as further evidence of the importance of information processing to the study of HCI. In light of this work, S-R compatibility is acknowledged as a factor influencing user stress and cognitive load (Shneiderman, 1992).

## 5.2 Directional Compatibility and Cursors in GUIs

In the present discussion, we will look at how S-R compatibility impacts user performance when pointing with various graphical cursor representations as they might be seen in typical WIMP-style GUIs. “Cursor representations” refer to the external graphical shape or appearance of cursors on a graphical display, and should not be confused with the definition



Figure 5.1: Graphical cursors as they appear in several popular GUI-based systems. The most popular pointing cursor in these GUIs is an arrow oriented toward the upper-left corner of the display. Other pointing cursors may not be explicitly shaped as an arrowhead but may convey directional cues nonetheless, such as the “pointing hand” and pen-shaped cursors in the figure. These cursor representations tend to vary little across systems and thus constitute a graphical standard that does not appear to have been extensively evaluated.

of internal mental representations provided earlier (Chapter 2). Most existing cursors for pointing include directional cues in the cursor itself that might implicitly affect the mapping between stimulus and response. Figure 5.1 presents example cursors from several popular GUI-based systems, noting how these cursors are represented with generally little variation. Little work currently exists with respect to the issue of cursor orientations for mouse input, and there has been no previous work to look at how the directional cues provided by pointing cursors could impact performance in more ubiquitous scenarios, such as distal pointing with a freehand pointer or surface-contact pointing with a pen.

A substantial amount of theoretical and applied evidence suggests that directional cues such as those found in pointing arrows can influence S-R compatibility. Experimental evidence by Kantowitz, Triggs, and Barnes (Kantowitz, Triggs, & Barnes, 1990) suggests that the perception of visual arrows in a task explicitly influences the directional compatibility of the task. The experiments by Worringham and Beringer (1989, 1998) involved physical controls with implicit directional cues such as levers, joysticks, and rotary knobs, demonstrating that these cues need not even be explicit to facilitate movement toward particular directions. Other work by Ristic, Friesen, and Kingstone (2002), and Tipples (2002) has found that arrowhead cues may be sufficient to induce certain kinds of reflexive shifts in visual attention.

Two experiments by Phillips, Meehan, and Triggs (2003; 2003) have directly evaluated arrowhead cursors on univariate (one-dimensional) mouse movements on desktop displays. Their experiments were explicitly designed to test for the presence or absence of S-R com-



---

patibility effects and their results suggest that the directional cues provided by an arrowhead cursor were sufficient to induce a systematic difference in user performance. However, contrary to most of the S-R compatibility literature, they found that arrowhead cursors predicted to be compatible with the direction of motion led to slower movements and less efficient cursor trajectories. Both of their experimental reports seem to imply that the illusory perception of target distance is primarily responsible for this phenomenon, but the authors also point out that other effects consistent with S-R compatibility were also observed, such as a systematic reduction in response latency for cursors cued in a compatible movement direction.

The two experiments by Phillips, Meehan, and Triggs are used as the motivation for the study of cursor orientations presented in this chapter. In both of the previous experiments, the experimental tasks were limited to movement along a single axis, and as such, they do not necessarily bear on the S-R compatibility effects associated with bivariate pointing movements. Those studies were also limited to the use of a mouse as input, and do not examine the impact of S-R compatibility effects with other kinds of input devices. It is unsafe to assume that because S-R compatibility effects exist with a mouse, such effects also exist with other kinds of input devices. Even if they do exist, what the magnitude of the effects might be is unclear. The psychological literature suggests that S-R compatibility is not only influenced by perceived visual cues but also by other factors such as hand posture and display-control orientation (Lippa, 1996). The present experiment described here independently builds upon the existing results of Phillips, Meehan, and Triggs (2003; 2003) by taking a representational perspective on S-R compatibility to provide further insight, and extending their investigation to bivariate pointing with a variety of input devices.

### **5.2.1 A History of Pointing Cursors**

Since graphical cursors are so integral to WIMP-based GUI systems, it may be argued that they constitute one of the most important elements of a graphical interface. The history of how pointing cursors in their various forms evolved over time is not particularly well-documented. This may be due to several factors including their near ubiquity on the desktop, which has led users and developers alike to take their presence for granted, or perhaps the engineering challenges posed by pointing devices such as mice were more important when cursors were first introduced, which meant that cursor appearance was

**CURSOR**

A small graphical symbol called a *cursor* is usually visible on your display screen. You can control the movements of the cursor on the screen with the *mouse*, or pointing device.


The shape of the cursor changes according to its location on the screen. This is to inform you what will happen if you press one of the buttons on the mouse, since the meanings of the buttons depend on the region of the screen you are in. The cursor's normal shape is a thin arrow pointing diagonally upward at an angle of about "eleven o'clock" like this: 

Figure 5.2: The definition of a graphical cursor excerpted from the Xerox Bravo document editing system reference manual, revision 5.1 (Xerox, 1980). The cursor arrow used in the Star and Alto systems developed at Xerox is indicated at the bottom of the figure, and has many of the same general features of the common upper-left pointing arrow used in the many WIMP-style GUIs that followed, such as the distinct straight edge along the left side of the arrow and the slanted edge along the right side.

---

“secondary” to the technical problems posed by the hardware used to drive them. The following discussion highlights some of the more important developments in the history of graphical cursors for pointing to provide an historical context and motivation for the cursor research presented here.

Cursors were likely first introduced as a means for enabling interaction with the earliest graphical displays. Sutherland’s Sketchpad system (1963) provided visual feedback in the form of a cross-like pattern not necessarily to support user input, but rather to facilitate the algorithm used to track the motion of a light pen. Light pen tracking was already in common use, as evidenced by Sutherland’s off-hand remark that the “availability of computer controlled display systems and particularly of light pen devices for manual input made it almost inevitable that computers would one day be involved in engineering drawing.” Engelbart’s prototype mouse (1963) provided a similar crosshair, which he called a “bug.” Unlike Sutherland, whose light pen did not strictly need visual feedback to support users, the relative-mode nature of the mouse required it. Thus, it appears that the relationship between cursors and pointing began years before any WIMP-based GUIs had ever been developed.

Further cursor development occurred at PARC, a division of Xerox at the time. The Xerox Alto system developed at PARC implemented the first prototype WIMP-style interface, which originally defined many of the interaction conventions used in subsequent systems. The mouse was chosen as the primary pointing device, meaning that cursors were again established as a necessary element of the user interface but in a context different from the ones envisioned by Sutherland and Engelbart. The cursor chosen for pointing in the Alto system was an arrow aimed toward the upper-left corner of the system display and this cursor eventually became the interface standard for most WIMP-based GUIs to follow (see Figure 5.2).

According to Butler Lampson (personal communication, August 2, 2004), who was involved in the development of the Bravo document editing component of the Alto, this cursor dates at least as far back as early versions of the system in September of 1974, and the choice of an arrow was probably motivated by at least two reasons. First, the designers wanted a shape that would point to a character rather than overlaying it. Second, pointing toward the upper-left seemed most natural given the relative orientation of the mouse with respect to the display (at least for right-handed mouse users). The particular graphical

representation chosen, with a straight edge along the left side and a slanted edge along the right side, may have been the result of technical constraints: this may have also been one of the few cursor representations that looked acceptable on the bi-level graphics hardware available at the time.

The Alto cursor had another characteristic that remains today: its use as a state indicator, often referred to as an *iconic cursor*. The Alto cursor changed shape according to what operations were invoked. The origin of icons is unclear. Tilbrook (1976) and Baecker (1979) implemented iconic cursors to represent system state in Newswhale, which was an experimental newspaper pagination system, around the same time the Alto was under development at PARC (personal communication, R. Baecker, December 21, 2004).

It may be somewhat surprising that the design choices made in the early days of GUI systems continue to be used in practice in more modern-day systems. Though most other aspects of WIMP-style GUIs have scaled upward with relative improvements in graphical display and input technology, the pointing cursor remains mostly unchanged from its original representation as an arrow pointing toward the upper-left. In many existing GUIs, pointing devices other than mice, such as pointers and pens, are being “retrofitted” with mouse-like input properties, such as its graphical cursor representations, even though there has been little research to indicate that elements such as the arrow cursor may be equally appropriate for these other kinds of input.

### 5.3 Hypotheses and Predictions

Existing evidence from previous experimental work in S-R compatibility makes a plausible argument for the presence of directional compatibility effects with graphical elements containing implicit directional cues. The directionally-oriented cursors used in many GUIs could have a measurable impact on user performance with respect to pointing movement and positioning. Based on the literature, the presence of S-R compatibility leads to three specific experimental hypotheses:

1. Graphical cursors that indicate direction will show performance differences consistent with those reported in the S-R compatibility literature, regardless of input method. Specifically, pointing movement with directional cursors aligned with the axis of movement should yield superior movement times and positioning performance. When di-

rectional cursors are not aligned with the direction of movement, movement times and positioning performance should be relatively worse.

2. Cursors that do not impose any possible directional cues, so-called *orientation-neutral* cursors, should yield the best movement times and positioning performance overall, regardless of input method. Since orientation-neutral cursors do not confer any specific benefits or penalties toward particular directions of movement, averaged movement time and positioning performance across a variety of movement directions should be better relative to those cursors that are cued for movement in one specific direction. One might also expect that those directionally-specialized cursors may *only* outperform the orientation-neutral cursor when moving in the specialized direction.
3. Because the S-R compatibility literature suggests that compatibility is influenced by hand posture and spatial organization, the degree to which varying cursor representations will affect movement and positioning performance is dependent on the input method. Based on the HCI literature with respect to input device movement performance, it is predicted that input with a freehand pointer (i.e. a laser pointer-like device) will yield the largest effects, mouse input will yield intermediate effects, and pen input will yield the smallest effects (Myers et al., 2002).

## 5.4 Experiment 1: Comparing Cursor Orientations

To test these predictions, a controlled experiment was conducted to observe user performance in a circular menu selection task with five different cursor representations and three different input devices: mouse, pointer, and pen. The circular menu selection task was chosen both because it permits measurement of pointing performance along bivariate directions of movement and because it resembles an earlier study of circular menu selection performance conducted by Callahan, Hopkins, Weiser, and Shneiderman (1988) and several previous experiments from cognitive psychology examining various aspects of S-R compatibility (Lippa & Adam, 2001; Cho & Proctor, 2001).

A completely within-subjects experimental design was used. Participants were instructed to complete three blocks each with 120 menu selection trials. Each block evaluated pointing performance with a different input device. Figure 5.3 illustrates the progression of

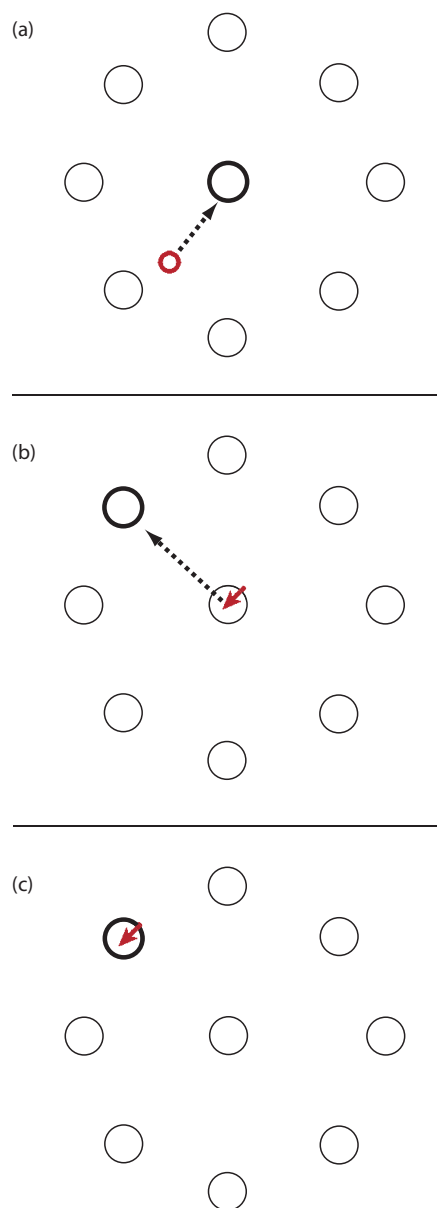


Figure 5.3: A typical trial as presented in this experiment. The dashed arrows indicate direction of movement and were not actually displayed to participants. (a) Participants started trials by selecting a circle located at the center of the display using the orientation-neutral circle cursor. (b) The cursor would either change to one of the four directional cursors, or stay the same for orientation-neutral trials. One of the eight circular menu positions would change colour, indicating where participants were expected to point. (c) Participants had been instructed to select the center of the highlighted item and to emphasize speed and accuracy equally.

---

a typical trial in the experiment. The order of trial block presentation was fully counterbalanced across subjects and the order of individual trials within blocks was fully randomized to minimize potential order effects. This randomization ensured that every block of presented trials had a unique ordering and that no two participants would ever see the trials presented in the same order.

### 5.4.1 Participants

Twelve participants were evaluated in this experiment. These participants were volunteers solicited from the undergraduate student population at the University of British Columbia in a manner consistent with the university's ethical guidelines for participant selection. Three participants were male and nine participants were female. Their ages ranged from 18 to 34 years. All participants had normal or corrected-to-normal vision. To minimize experimental bias due to handedness, all of the participants were required to be right-handed. Participants were asked prior to the beginning of a session what their preferred handedness was and those participants stating a preference for left-handedness were excused from the study. It was determined that this would be sufficient for the selection process and that no further handedness tests would be necessary.

### 5.4.2 Apparatus

Figures 5.4, 5.5, and 5.6 present the apparatus and input devices used in this experiment. Each block of experimental trials used a different input and display device for menu item selection. A PC workstation (dual-processor Pentium 4 3.0 GHz, 1 GB of RAM running Microsoft Windows XP Professional Edition, and an nVidia Quadro 4 video card) was used to present trials to participants in the blocks involving mouse and pointer input. A Compaq TC1000 Tablet PC (running Microsoft Windows XP Tablet PC Edition) was used to present trials in the block involving pen input. The experiment software was written using the Java SDK, version 1.5.0 with hardware-accelerated OpenGL support. Movement times were collected using a native high-resolution timer based on the CPU clock rate. This effectively permitted the capture of movement times with a sampling resolution of approximately one millisecond. In all conditions, a constant level of room illumination was maintained.

In the block of trials using mouse input (the "mouse condition"), a 17-inch LCD flat-panel display with a resolution of 1024x768 pixels was used (see Figure 5.4). A Logitech

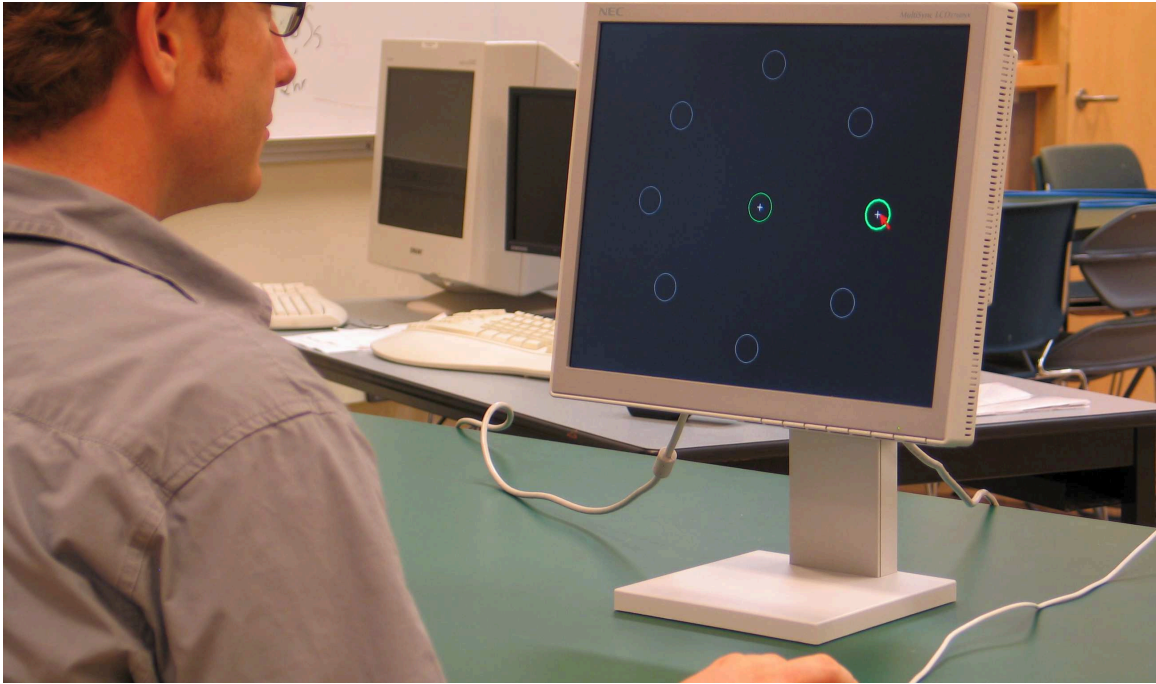


Figure 5.4: The mouse condition apparatus used in this experiment. Participants used a mouse to point at and select highlighted targets on a desktop-size display.

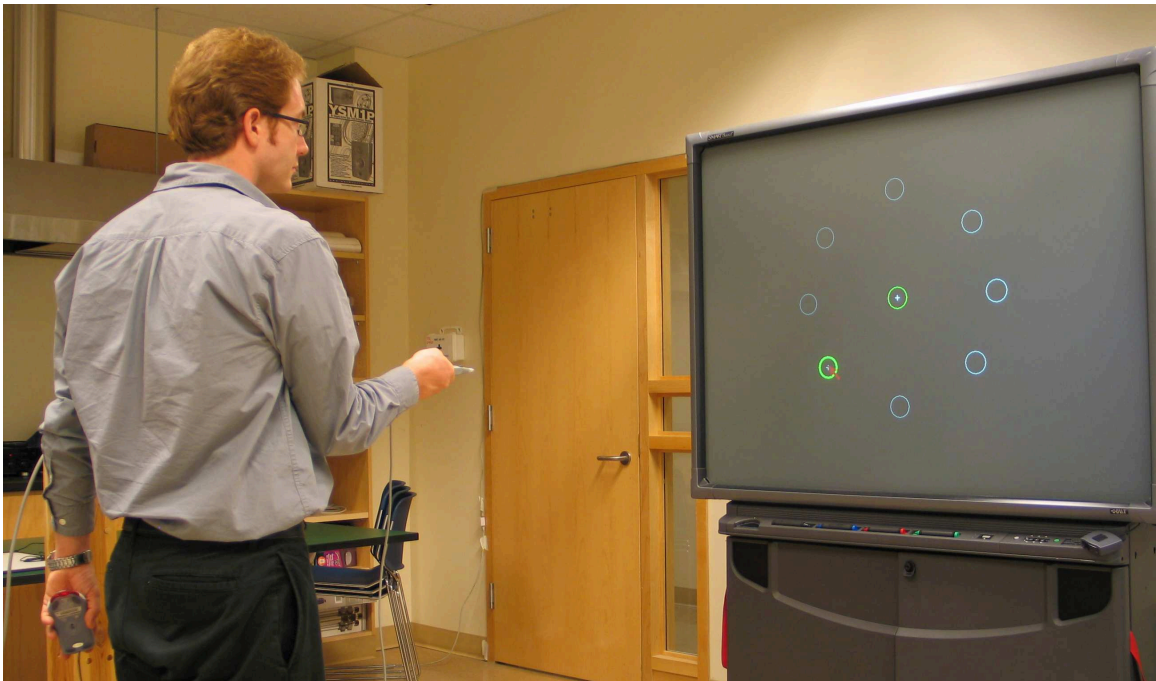


Figure 5.5: The pointer condition apparatus used in this experiment. Participants used a Polhemus Fastrak-based 3D pointer simulating a laser pointer to point at highlighted targets on a large-screen SMART Board display. Items were selected by pressing the left button on a two-button mouse held in the left hand.





Figure 5.6: The pen condition apparatus used in this experiment. Participants used a Tablet PC and its accompanying pen to select highlighted targets on this portable display. Selection was indicated by directly tapping highlighted items with the pen.

optical Wheel Mouse was used as the input device. To control for mouse gain (sometimes referred to in the literature as the control-display ratio), a hand-display movement ratio of 1:1 was used so that the mouse moved the same distance in hand space as it did in display space. This was done intentionally to permit a fair comparison of compatibility effects across the different input devices evaluated in this experiment.

In the pointer condition, a 66-inch LCD rear-projection SMART Board 3000i display was used (see Figure 5.5). The display was set to a resolution of 1024x768 pixels. Although SMART Boards are generally used for touch interaction, its touch functionality was not used in this experiment; it was only used as a physically-large graphical display. The 3D pointer of an electromagnetic Polhemus Fastrak was held in the right hand of participants while a two-button mouse was held in the left hand. The Fastrak transmitter and 3D pointer were kept away from metallic interference and the tracking system was calibrated and registered before each experiment session. Pointing was accomplished with the 3D pointer. Items were selected by pressing one of the mouse buttons.

In the pen condition, a Compaq TC1000 Tablet PC with a resolution of 1024x768 pixels

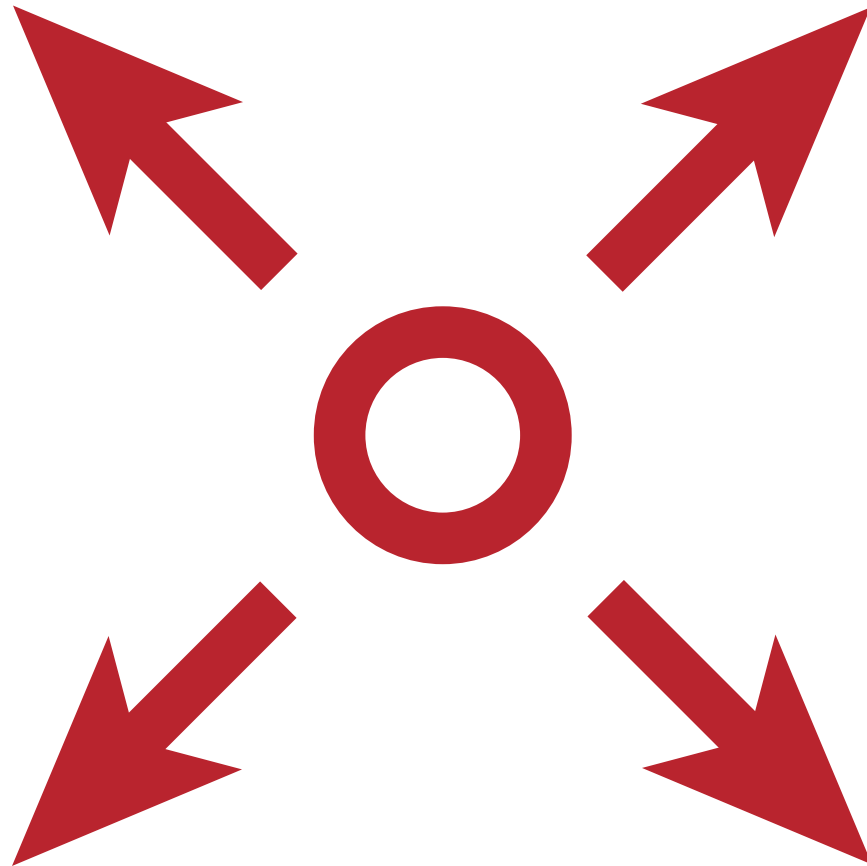


Figure 5.7: The five cursor representations evaluated in this experiment. The four arrow cursors were oriented at 45 degree angles toward the corners of a display. The orientation-neutral cursor was a hollow circle with no implicit directional cues toward any particular direction of movement.

was used (see Figure 5.6). The Tablet PC operated in landscape (wide) format to provide an aspect orientation consistent with the other two display conditions and the display was inclined at an angle of 45 degrees. The Tablet PC pen and display used for input were calibrated to correct for viewing parallax before each experiment session by using the calibration software provided by Windows XP.

### 5.4.3 Procedure

Participants took part in a single session approximately forty minutes in duration during which they completed all three blocks of trials. During sessions, a researcher was present at all times. Participants were given five-minute breaks between each block of trials.

In the mouse condition, participants were seated at a distance of 50 cm from the LCD

---

monitor. In the pointer condition, participants stood upright at a distance of 200 cm from the SMART Board display. In the pen condition, participants were seated at a distance of 30 cm from the Tablet PC. Trials in each block consisted of a single pointing movement with a given input device, starting from the center of a display and moving to one of eight equally-distant menu items with one of five possible cursor representations. Figure 5.7 illustrates the graphical cursors used in this experiment: arrows pointing toward the (1) upper-left, (2) upper-right, (3) lower-left, (4) lower-right of a display, and (5) an orientation-neutral circle. The arrow cursors were specifically designed to be oriented at 45 degree angles to allow for a systematic comparison of directional compatibility.

All of the visual elements in the experiment were rendered to subtend the same visual angle from the perspective of the participant across each of the different display and subject viewing distances, consistent with experimental practice in cognitive psychology. All cursors, including the orientation-neutral circle cursor, subtended a visual angle of two degrees in all of the display conditions. Target menu items were eight circles, four degrees of visual angle in diameter and ten degrees distant from the center of the display. These targets were equally spaced in a circular fashion around the center target to form a circular menu. Each combination of trial parameters was repeated three times, yielding a total of  $5 \text{ cursor orientations} \times 8 \text{ menu positions} \times 3 \text{ input devices} \times 3 \text{ repetitions} = 360 \text{ trials}$  per participant. Thus, each trial block consisted of a complete set of 120 trials using one particular input device.

### **Mouse Condition**

In the block of mouse input trials, participants were provided with a Logitech optical Wheel Mouse, which was placed on a desktop (table) directly in front of them. Participants pointed at onscreen items by moving the presented visual cursor to the graphical items using the mouse and then clicking the left mouse button to indicate selection.

### **Pointer Condition**

In the block of pointer trials, participants were instructed to hold a Polhemus Fastrak 3D pointer (freehand stylus) in their right hand while holding a push button mouse in their left hand. Participants pointed at onscreen items by directly aiming the 3D pointer at graphical items and clicking a button on the mouse held in their left hand. This two-handed pointer-

and-mouse approach permitted the capture of participant performance data without the potential confounding errors associated with the “pen-drop” phenomenon often associated with pointer input (Myers et al., 2002).

### **Pen Condition**

In the block of pen input trials, participants used the tablet-sensitive pen of the Tablet PC for item selection. Although the Tablet PC was situated on a table, participants were instructed to hold the Tablet PC with their left hands for additional stability. The pen was held in the participants’ right hands and they pointed at onscreen items by directly tapping at graphical items to indicate selection.

### **Training**

All participants received a minimum of twenty practice trials prior to each block of experimental trials. Practice trials consisted of trials presented in the same manner as true experimental trials. In the practice trials, cursor orientations and items for selection were chosen at random. Practice trials were presented until both the participant and researcher were satisfied with the participant’s ability to complete the task as instructed.

#### **5.4.4 Data Analysis**

During the experiment, two primary quantitative measures were collected. Total movement time was defined as the period of time from trial initiation to the point at which a highlighted item was selected, measured in milliseconds (ms). Movement precision was measured in two different ways. First, it was measured as a standard root-mean-square (RMS) index from the center of the intended target menu position of every trial. Second, movement precision was measured in terms of effective target width, which is sometimes computed when analyzing 2D movement performance with Fitts’s Law, using the derived RMS values as a measure of pointing variance (MacKenzie & Buxton, 1992).

Collected participant data were analyzed statistically in a manner consistent with the outlined experimental hypotheses. To determine whether the participant data supported the claims described by the experimental hypotheses, a factorial analysis of variance with repeated measures (ANOVA-RM) over three independent factors of input device, menu position, and cursor type for the dependent measures of movement time, RMS index, and

---

effective target width. Sphericity was evaluated using Mauchly’s test and visual observation of the data. Mauchly’s test suggested that some measures in some cells might have violated the assumption of sphericity ( $p$ -values ranged from .004 to .523 across cells and measures), so a Greenhouse-Geisser correction was applied to adjust the estimated degrees of freedom in all relevant statistical analyses. In the statistics that follow, all relevant degrees of freedom are reported as adjusted by the Greenhouse-Geisser correction.

If changes in cursor representation had an impact on movement performance, these would be observed in the ANOVA-RM as statistically significant ( $p < .05$ ) two-way interaction effects between menu position and cursor type across all three dependent measures. To assess the overall impact of movement performance for different cursors along all directions of movement, which is important when considering general visual feedback solutions for GUIs using pointing as a means for interaction, corresponding main effects for cursor type were also of interest. Averaged movement times, RMS indices, and effective target widths were computed across all eight menu positions for each cursor type as well. These averaged movement performance measures were subjected to an additional two-way ANOVA across input device and cursor type.

#### 5.4.5 Movement Time Performance

Consistent with the experimental hypotheses, the factorial three-way ANOVA-RM found a significant two-way interaction between menu position and cursor type [ $F(2.464, 27.105) = 3.875, p = .033$ ]. No higher-order interaction effect was observed, suggesting that the way in which a cursor was represented had an impact on movement time performance toward particular directions of movement. A corresponding two-way interaction between input device and cursor type was also statistically significant [ $F(2.751, 30.261) = 3.578, p = .028$ ]. This result supports the hypothesis that the impact of different cursor types on movement time differs across input devices.

Participant performance was compared in those trial situations where cursor type and menu position were compatible (i.e. when an upper-left cursor was used to point at the menu item in the upper-left position and so on) versus the averaged movement time for that particular menu position, regardless of cursor type. This comparison was done to assess the difference in effect size between compatible mappings and averaged “expected” movement performance. In all of the compatible situations, movement times yielded consistent average

improvements of approximately 35 ms for mouse input (approximately 3.3 percent faster than average), 65 ms for pointer input (approximately 5.5 percent faster), and 17 ms for pen input (approximately 1.5 percent faster).

An identical comparison was done to measure the difference in effect size between purely incompatible mappings (i.e. when an upper-left cursor was used to point at the menu item in the lower-right direction and so on). In these incompatible situations, movement times yielded consistent effects although these performance penalties were relatively smaller. Mouse input was observed to be about 18 ms slower (approximately 1.7 percent slower), pointer input was 43 ms slower (approximately 3.7 percent slower), and pen input was about 9 ms slower (approximately 0.8 percent slower) on average.

Also consistent with the experimental hypotheses, the ANOVA-RM revealed a secondary main effect of cursor type, suggesting that by simply varying the cursor representation that was used for movement alone, it was possible to manipulate movement time performance [ $F(2.279, 25.072) = 5.545, p = .008$ ]. No other main effects were observed for either input device [ $F(1.710, 18.811) = 2.097, p = .155$ ] or menu position [ $F(3.061, 33.671) = 2.544, p = .072$ ].

Figures 5.8, 5.9, and 5.10 present bar graphs of averaged movement time across all tested movement directions, by cursor type for each input device. The secondary two-way ANOVA for input device and cursor type for movement times averaged across all eight menu positions yielded a two-way interaction effect across both independent factors, suggesting that some cursor types performed better than others overall and that the degree to which these cursors differed also depended upon input type. The orientation-neutral circle cursor was found to have the best averaged performance overall and this was consistent across all three tested input devices. The cursor oriented toward the lower-right of the displays consistently had the worst averaged performance overall for all input devices. This difference was most apparent with pointer input, which demonstrated a movement time difference of 157 ms between the orientation-neutral circle cursor and the lower-right arrow cursor. These differences were substantially smaller for mouse and pen input, with movement time differences of 59 ms and 27 ms respectively.

The orientation-neutral circle was found to consistently outperform the traditional upper-left arrow cursor that represents the orientation used in many existing GUI systems. Performance differences between the upper-left arrow cursor and orientation-neutral circle

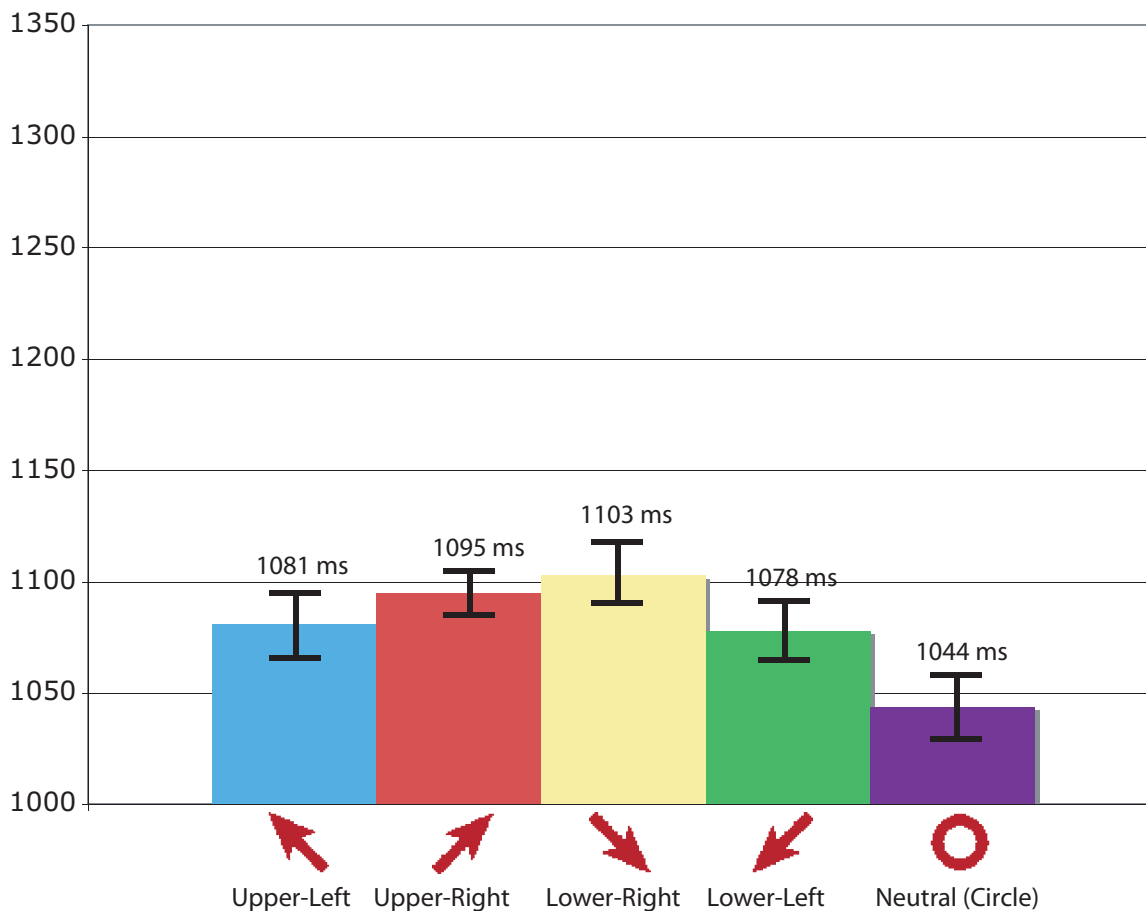


Figure 5.8: Bar graph broken down by cursor type for averaged mouse movement times across the eight menu positions in this experiment.

were most apparent with pointer input, where movement times differed by 127 ms. With mouse input, these movement times differed between both cursors by approximately 37 ms. With pen input, these movement times differed by approximately 17 ms. Although small, these differences are measurable improvements in movement time performance, suggesting that there may be situations in which an orientation-neutral cursor might be beneficial over the standard upper-left arrow cursor typically employed.

A final analysis of movement time was conducted to examine the average movement times across all eight tested menu positions and all five cursor types. This analysis was done to look at the overall impact of device performance on the observed cursor effects, independent of S-R compatibility. A one-way ANOVA across all three input devices found that movement time performance was significantly different across mouse, pointer, and pen input [ $F(1.682, 18.503) = 62.977, p < .001$ ], consistent with previous research reported in

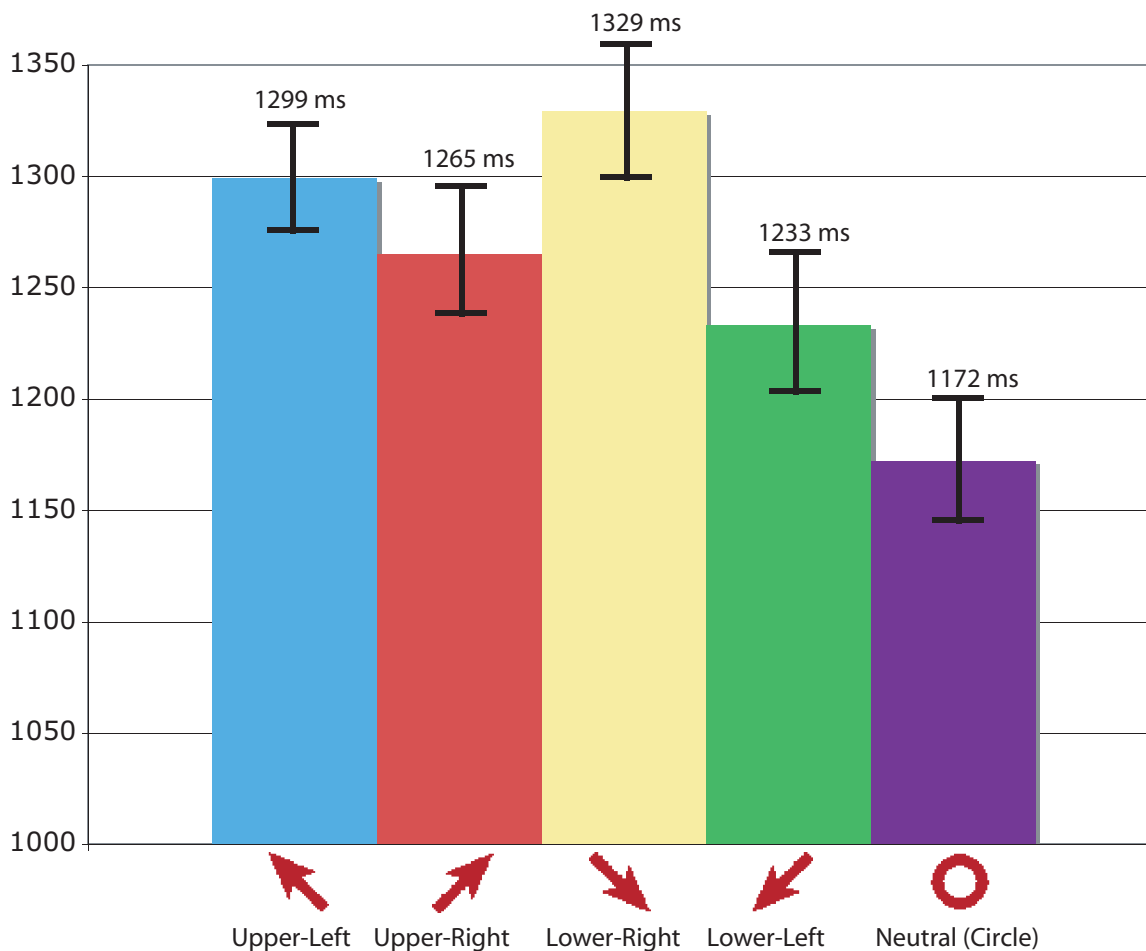


Figure 5.9: Bar graph broken down by cursor type for averaged pointer movement times across the eight menu positions in this experiment.

the HCI literature. Average mouse movement times were 1075 ms in duration, pointer movement times were 1193 ms in duration, and average pen movement times were 1081 ms in duration. Thus, pointer input appeared to be the least efficient, having the longest movement times, while mouse and pen input were relatively similar in performance.

#### 5.4.6 Precision and Accuracy

Results for measures of movement precision were mostly consistent with those found for movement time. While the primary factorial ANOVA-RM did not find any higher-order interaction effects (menu position by cursor type: [ $F(7.223, 79.452) = 1.427, p = .205$ ]), it observed a statistically significant main effect for input device on measured RMS index and effective target width [ $F(1.197, 13.162) = 1840.995, p < .001$ ]. RMS index and effective



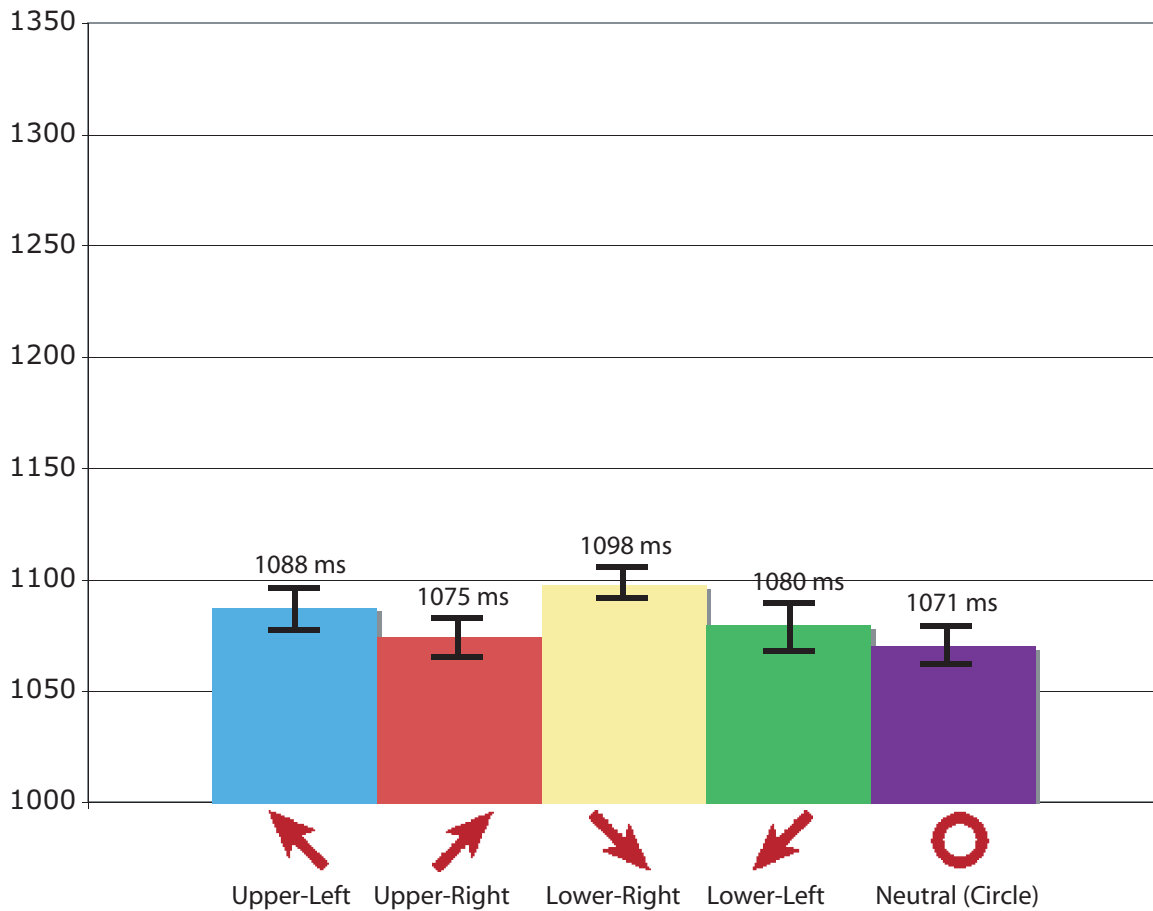


Figure 5.10: Bar graph broken down by cursor type for averaged pen movement times across the eight menu positions in this experiment.

target width share the same F-values since effective target width was computed as a function of RMS. Both mouse and pen input were found to have roughly comparable movement precision performance, but pointer input was substantially worse. Pointer interaction, in particular, may have suffered because of movement effects such as hand jitter due to lack of a stable surface for pointing.

The actual target widths of target menu positions were four degrees of visual angle in size. The effective target widths for these menu positions remained roughly the same for mouse and pen input. However, the computed effective target width for pointer input was nearly eight times larger (i.e. the spread of pointing errors was much larger when pointing with a pointer than with a mouse or pen). This is an indication that participants likely spent more time making movement corrections with pointer input than they did with either mouse or pen input.

The ANOVA-RM also found a main effect for cursor type on measured RMS and effective target width [ $F(2.586, 28.441) = 3.176, p = .045$  for both measures], but the overall influence of these effects were much smaller (partial  $\eta^2 = .224$ ) compared to that of the main effect for input device (partial  $\eta^2 = .994$ ). Thus, the experiment was unable to uncover any predicted relationship between cursor type and movement direction as postulated in the experimental hypotheses, although it may be the case that any movement positioning effects due to S-R compatibility might have been masked by input device performance as evidenced by the large differences in the partial  $\eta^2$  values in the tested factors.

With respect to cursor type, computed effective target widths were smaller than actual target widths when using the orientation-neutral circle cursor and upper-right arrow cursor. However, these improvements were relatively small, demonstrating a width decrease of 0.3 degrees for the orientation-neutral cursor and 0.15 degrees for the upper-right arrow cursor. All of the other cursor representations yielded differences between effective target width and actual target width of less than 0.1 degrees, which is not large enough to be considered practical or meaningful. Because effective target width was directly derived from the RMS measure, these agreed with values for movement precision as measured by RMS index. A review of RMS index values yielded very small, but consistent, differences in favor of the orientation-neutral circle cursor compared to all of the other cursor representations.

#### 5.4.7 Other Observations

Participants were asked at the end of each session to provide a subjective, qualitative preference for each of the cursor and input types tested. Some participants stated a clear preference for the orientation-neutral circle cursor while others felt that the circle was “easier” to use, though none of the participants were able to provide a good explanation for why they felt this way. Two of the participants remarked that they found it surprising that orientation-neutral cursors were not used more often in desktop GUI settings. These subjective preferences and observations are well-supported by the empirical data obtained in this experiment.

With respect to input device, several of the participants expressed a sense of novelty with respect to using pointer and pen input. All of the participants claimed to only have substantial experience with using a mouse for pointing, so this particular observation may not be surprising. One participant remarked on the use of the pointer on a large screen as

---

interesting and as having potential for games and entertainment, and several other participants felt that they would probably like to be able to use it for short amounts of time when interacting with a large screen. However, all participants agreed that the use of the pointer would probably lead to unnecessary fatigue during extended use.

## 5.5 Discussion of Experiment 1

This experiment demonstrates that *S-R compatibility and the integration of visual information for perception and action can systematically influence user performance*, even in the case of a simple graphical element like a visual cursor. The influence of S-R compatibility was observed to be dependent on the input and display configuration being used. Impact was greatest when interacting with a pointer on a large screen while impact was minimized when interacting with a pen on a portable display. These results overall suggest that the kinds of visual cues used in existing and future GUI systems need to be considered carefully by designers. Moreover, knowing that many elements can influence S-R compatibility, including cursor representation, hand posture, and user orientation, is an important step along the way toward understanding the basic elements that constitute effective interaction with GUI systems.

The experiment and results presented here open up the possibility for further study and additional research. Although this experiment did not find any specific S-R compatibility effects for movement precision, it remains unclear whether different directional cursors can influence movement precision. The previous experiments by Phillips, Triggs, and Meehan (2003; 2003) certainly imply that this is the case. The inability to find any statistically significant differences may be the result of other factors such as input device and movement amplitude (Phillips, Triggs, and Meehan suggest that the influence of compatibility on movement precision is a function of movement distance). Additionally, this particular experimental design was chosen to minimize bias with respect to handedness and learning effects such as contextual interference. Future experiments may look at the aspect of handedness with respect to S-R compatibility more carefully. It would be interesting to determine if any “mirrored” compatibility effects exist, as might be suggested by the S-R compatibility literature with respect to the relationship between display and control.

Another interesting follow-up would involve manipulation of different kinds of orientation-

neutral cursors. Though a circle was used in this experiment, other studies involving orientation-neutral cursors use crosshairs or other similar representations, possibly because these are the most prevalent non-directional cursor representations in use. It would be valuable to determine whether such cursors yielded equivalent performance to the circle cursor used here, or if other implicit cues in different orientation-neutral cursors might further influence the S-R compatibility of user input.

### 5.5.1 Implications for User Interface Design

Based on the present experiment, the following implications for user interface design are suggested. These implications provide evidence that a representational approach to HCI can not only be theoretically interesting, but can also be helpful in the practical design and evaluation of interactive systems.

1. *Where appropriate, consider the use of orientation-neutral visual cursors and be careful with the use of visual cursors that may include directional cues.*

Using a cursor representation free of directional bias was observed in this experiment to provide faster movement performance and better movement precision. This finding could be helpful in GUI systems that are organized in such a way that interactive elements are located all over the display (such as in existing WIMP-based systems). Consistent with the literature looking at judgments of S-R compatibility, designers should not assume that a directional cursor might not measurably affect performance because evidence exists that orientation-neutral cursors may better facilitate user performance.

2. *In situations where many interactive elements may be clustered in one particular area of the display (i.e. palettes or toolbars), choosing a cursor tailored to the direction of movement could be helpful.*

There may be particular instances where the inclusion of directional cues may be desirable. In this experiment, it was found that cursors oriented in particular directions facilitated movement in those directions. Choosing a cursor tailored to one specific direction might be helpful when users are expected to point along a particular axis of movement fairly frequently. Furthermore, such cues might be helpful as a visual aid to encourage users to make movements in particular directions (as might be the case

---

when interactive shortcuts or visual mnemonics are employed to facilitate expert user performance).

3. *Where movement precision and positioning are important, consider using an input device that rests on a stable surface or consider using a touch-based input method such as pen input.*

The experimental results were found to be consistent with previous work investigating user performance with different kinds of input. Pen input was shown to have the least “susceptibility” to predicted S-R compatibility effects with respect to cursor type, which may be because such visual feedback was unnecessary for the pointing movements required in the experiment. Other visual feedback for pen input was already present, including sight of both pen and limb. In instances where pen or even mouse input may not be viable, S-R compatibility becomes more important. The results of this experiment suggest that at least one way to increase user performance with pointer input is to provide compatible visual cues that facilitate movement performance.

4. *Do not assume that it is easy to design a user interface that has compatible S-R mappings and be aware that visual elements can inadvertently affect the compatibility of user input.*

This experiment demonstrates that the standard pointing cursor used in many existing GUI systems can be replaced by other, more compatible elements even though many designers would assume that choice of cursor representation would make no difference whatsoever, and that even if it does, that the “tried-and-true” upper-left arrow pointing cursor would be best. The results of this experiment suggest that designers need to be willing to look beyond their own intuition when building an effective, efficient user interface because many performance-altering phenomena such as S-R compatibility are not readily apparent without systematic study.



## Chapter 6

# Specialization of the Upper and Lower Visual Fields <sup>1</sup>

The separation of visual information for perception and action can be observed in the evolution of the human eye and the functional specialization of the upper and lower visual fields (UVF and LVF). Unlike the integration of visual information as seen in the phenomenon of S-R compatibility discussed in the previous chapter, the UVF and LVF characterize a perspective on mental representations of visual space whereby vision for perception and vision for action are separate and unique representations of visual space. Although it may seem paradoxical that visual information could be simultaneously integral and separable, we only need to refer back to the computational theory of mind to understand how this is possible. Marr (1982) in particular points out that much of the difficulty inherent in solving a problem (or carrying out a task) is a function of how the information available to solve the problem is represented. When information is represented in a particular way, certain problems may be easier to solve at the expense of making other problems more difficult to solve. Thus, any complex system that must be capable of carrying out *multiple* tasks may have to use multiple representations of the same information to efficiently carry out *particular* tasks.

In the case of human vision, there is substantial evidence that supports the processing of visual information both as an integrated entity and as a separated entity. For some kinds of visual operations, it may be more efficient to represent visual information in an integrated mapping, as characterized by the situations in which phenomena like S-R compatibility have been observed to appear. As we will see, there may also have been factors in the evolutionary history of humans that could have led to the development of separable representations

---

<sup>1</sup>A version of this chapter has been published. Po, B. A., Fisher, B. D., and Booth, K. S. (2004). Mouse and touchscreen selection in the upper and lower visual fields. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 359-366.

of visual information for perception and action as an evolutionary advantage. Based on current experimental evidence, the consequences of these evolutionary factors continue to exist today, even though the world that people inhabit is substantially different from the one that existed many thousands of years ago. Such evolutionary choices may be even more important now, as the visual world is arguably more complex today than it has been at any other point in human history. Identifying the impact of separate representations for perception and action could be important to fully understanding user performance, and the functional specialization of the UVF and LVF could have important implications for user interface design.

## **6.1 Background and Related Work**

Archaeological evidence suggests that the anatomical characteristics of the modern human started to appear some 170,000 years ago (Wynn, 2002). The development of vision as a sensory modality predates this, and there is phylogenetic evidence suggesting that all biological vertebrates that now possess vision are descended from organisms who developed vision approximately 540 million years ago (Nilsson, 1996). It is postulated that vision probably first evolved to facilitate movement and that other characteristics typically associated with the sense of sight developed much later. It is further believed that it is this distinction that led to the development of separate representations of visual space for perception and action. This ecological argument implies that the neural pathways that process visual information for movement have had a longer evolutionary history than those pathways that process visual information for perceptual functions.

### **6.1.1 The Functionally Specialized Eye**

Some ecological theories suggest that four important changes in the visual environment of primates are responsible for the division between vision for perception and vision for action (Previc, 1990). First, primate vision underwent a tremendous increase in the optical resolution of the eye (Polyak, 1957). Second, primates became more reliant on coloured fruits as a food source, made partially possible by the evolution of spectrally-sensitive cone pigments (Snodderly, 1979). Third, facial expression became an important instrument of emotional expression and social communication (Allman, 1977). Fourth, the primate visual



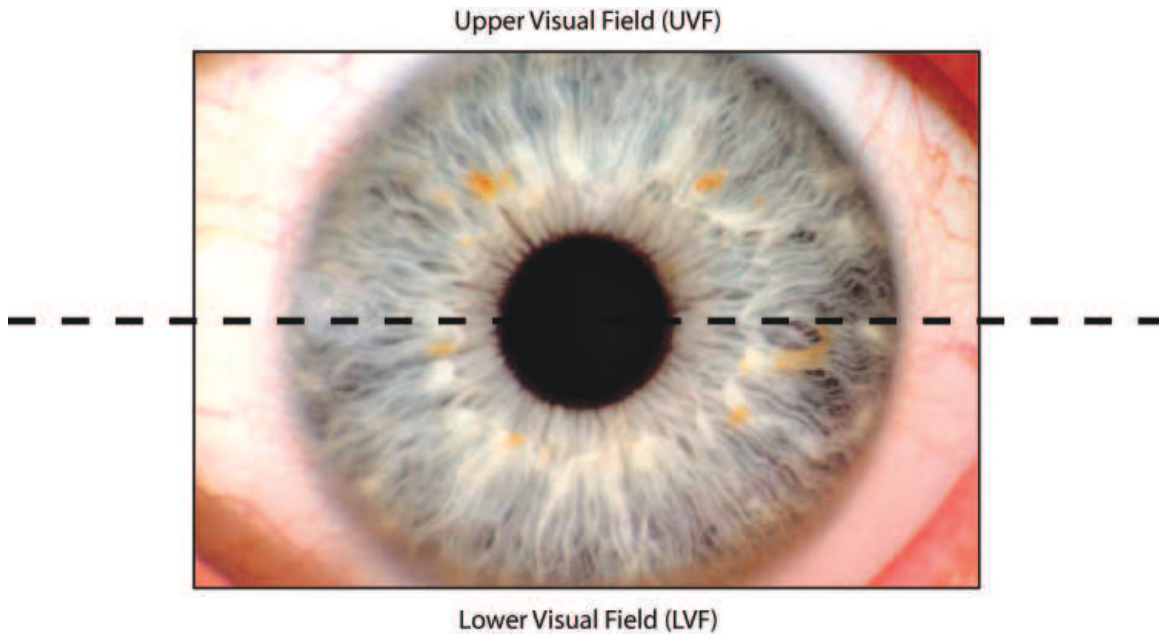


Figure 6.1: An illustration of the upper and lower visual fields (UVF/LVF). The UVF corresponds to the upper half of what is seen while the LVF corresponds to the lower half.

system evolved the capacity to make voluntary saccadic eye movements independent of head movements (Previc, 1990). Thus, primates eventually developed the need for a visual system capable of supporting efficient *perceptual* functions and the mechanisms to meet those needs.

Two other critical changes occurred concurrently with these perceptual developments, leading to the need for a visual system simultaneously capable of supporting efficient *visuomotor* functions. First, primates had developed an increased body size and a sitting or partially erect posture became regular behavioural practice. This resulted in an elevation of the eyes relative to the rest of the body, altering the role of the hands and arms for primarily manipulative functions rather than for postural support (Osman Hill, 1972). Second, changes in the shape of the hand led to an increased capacity for sophisticated reaching behaviours in higher primates (Bishop, 1962). This permitted higher primates to perform finer and more skilled motor functions than were previously possible.

These evolutionary theories further suggest that these changes to the visual environment correspond to the development of differential visual processing dependent on where visual stimuli were detected. Since many of the perceptual tasks that occurred also happened at a distance, these were more likely to occur in the upper half of what was seen. Similarly,

<b>Upper Visual Field (UVF)</b>	Specialized for <i>perceptual</i> activities in far, or extrapersonal, space. Superior performance for activities like visual search and object recognition. Longer feature persistence and better discrimination of facial features.
<b>Lower Visual Field (LVF)</b>	Specialized for <i>visually-guided motor</i> activities in near, or peripersonal, space. Greater attentional resolution and acuity for fine detail. Better temporal change discrimination.

Table 6.1: A summary of the functional differences between the UVF and LVF based on evidence from the psychological literature.

because the hands and arms became necessarily situated below the eyes, many visuomotor tasks became more likely to occur in the lower half of what was seen. This apparent division in visual tasks led to a major functional division between the vertical hemifields of the human eye. In modern psychological terms, the upper visual field (UVF) corresponds to the upper half of the perceived visual world while the lower visual field (LVF) corresponds to the lower half of the perceived visual world. Figure 6.1 illustrates this division and Table 6.1 summarizes the most important functional differences between these two visual fields.

The UVF is believed to be functionally specialized for activities in *far*, or *extrapersonal* space, which refers to activities that occur outside of reaching distance. This is an interpretation that has arisen from psychological evidence suggesting that the UVF supports perceptual activities like visual search and object recognition. Likewise, the LVF is believed to be functionally specialized for activities in *near* or *peripersonal* space, which refers to activities that occur within reaching distance. This has arisen from psychological evidence suggesting the LVF supports visually-guided motor activities.

Despite the use of the term “functional specialization” to refer to the distinction between the UVF and LVF, it is not the case that particular functions can only occur in one or the other visual field. Rather, the cumulative evidence supporting the presence of specialized functionality in the UVF and LVF indicates that *both* visual fields are capable of performing the same tasks, but one of the two fields has superior task performance relative to the other.

Relevant evidence to support the UVF and LVF distinction falls into three primary categories: reaction time performance, visual attention, and neuroanatomical features. Other evidence includes experimental work in the areas of motion perception and visual evoked

---

potentials (Previc, 1990). Skrandies (1987) and Previc (1990) provide substantial reviews of this evidence, some of which is discussed here. The evidence cited in their reviews is further supported by additional experimental evidence reported in the years following their publications, some of which is also discussed here. All of this evidence supports the notion that the UVF and LVF separate visual information into two distinct mental representations during processing, with the UVF supporting a representation of visual space for perception and the LVF supporting a representation of visual space for action.

### 6.1.2 Reaction Time Performance

Differences in reaction time are some of the best studied and most reliable evidence that differences exist between the UVF and LVF. Woodworth (1938) cites evidence suggesting that reaction time latencies for most visual stimuli are shorter when such stimuli are presented to the LVF. Payne (1967) subsequently characterized these differences as being eight to ten milliseconds shorter at the vertical division between the UVF and LVF but more than twenty milliseconds shorter when well inside the LVF. Gawryszewski et al. (1987) and Rizzolatti et al. (1987) have replicated these results, showing that such reaction time differences arise at least under valid or neutral attentional cueing.

Tychsen and Lisberger (1986) report a consistent asymmetry in eye movement accelerations in the UVF and LVF. They found that target pursuit had greater acceleration for both upward and downward target motion when targets moved into the LVF but that saccadic eye movements were faster to static targets when such stimuli were presented in the UVF. This is consistent with a review of experimental evidence by Heywood and Churcher (1980), which found that most published studies reported superior saccadic eye movement performance for visual items presented in the UVF.

Christman (1993) used a hierarchical letter stimuli task to demonstrate that global (group) percepts were detected more quickly when they were presented to the LVF while local features were detected more quickly when they were presented to the UVF. Rubin, Nakayama, and Shapley (1996) found an LVF advantage for the perception of illusory contours, consistent with the results from the earlier Christman study. Niebauer and Christman (1998) used a comparison of coordinate and categorical judgments toward the same stimuli and found that reaction times were faster to LVF stimuli for coordinate judgments but not for categorical ones, consistent with the interpretation that the LVF is specialized

for visually-guided movements. A psychophysical pointing task by Danckert and Goodale (2001) found that direct, physical pointing toward physical (cardboard) targets was faster and more accurate when such items were perceived in the LVF.

### 6.1.3 Visual Attention

The experimental work surrounding the topic of visual attention has uncovered further evidence of vertical asymmetries in the visual field. Gawryszewski et al. (1987) investigated (physical) manual responses to perceived visual targets, suggesting that a three-dimensional cubic framework for spatial visual attention might exist. In their work, fundamental divisions in movement performance occurred along a depth (near-far) axis, a lateral (left-right) axis, and a vertical (upper-lower) axis corresponding to the UVF and LVF. Thus, they concluded that because there is an implicit relationship between spatial attention and visuomotor coordination with a bias toward the LVF, the LVF may have been functionally specialized to support the accurate monitoring of the trajectory of a reaching hand. More recent work by Intriligator and Cavanagh (2001) found that the spatial resolution of visual attention, or the capacity to individuate items in a presented group of items, was asymmetric in the UVF and LVF. They found that the “critical spacing” for individuation was larger in the UVF than in the LVF, consistent with the notion that finer individuation may be an important feature for highly skilled motor movements but may be less important for perceptual tasks in extrapersonal space.

Several experimental studies have found that there is a bias toward visual search tasks in the UVF, consistent with the hypothesis that the UVF is specialized for perceptual activities. Jeannerod et al. (1968) found that visual search usually initiates in the UVF, possibly explaining why targets in the UVF are more frequently identified in briefly presented displays (Goldstein & Babkoff, 2001). Chedru et al. (1973) also found that visual searches were disproportionately biased toward an examination of the UVF, consistent with the finding that while performing a search of a visual display in memory, subjects typically elevate their eyes (Kinsbourne, 1972). Other work by Piaget (1969) found that a vertical line intersecting a horizontal line perceived in the UVF appeared shorter than in the LVF and he argued that this was the result of attentional focus being “shifted” toward the UVF. More recently, Heider and Groner (1997) showed in a comparison of the vertical visual fields that detected colour features were more persistent in the UVF relative to the LVF.

---

Patients with spatial neglect syndrome (a condition often caused by a stroke, characterized by a measurable inability to perceive and attend to objects in particular regions of the visual field) also provide supporting evidence for a distinction between the UVF and LVF. Rubens (1985) found that all eighteen patients diagnosed with spatial neglect in his study had more pronounced symptoms in the left-hand side of the LVF relative to the rest of the visual field and that thirteen of these patients had spatial neglect that was *limited* to the LVF. Such evidence is consistent with the experimental ablation work on monkeys conducted by Rizzolatti et al. (1985) who found that selective induction of spatial neglect in extrapersonal or peripersonal space was possible by creating lesions in regions of the brain directly related to processing of visual information in the UVF and LVF.

#### 6.1.4 Neuroanatomical Features

Anatomical evidence indicates that the distribution of ganglion cells in the eye across the UVF and LVF is non-uniform. Two studies by Curcio et. al (1987) and Curcio and Allen (1990) indicate that there are greater concentrations of ganglion cells in the LVF, perhaps anatomically consistent with the findings of Intriligator and Cavanagh (2001), suggesting a greater spatial resolution for visual attention in the LVF. The greater cell density in the LVF is also consistent from an ecological perspective. What occurs in the LVF is likely of particular interest immediately, while what occurs in the UVF may be of interest only sometime in the future (although perhaps on the order of several seconds later). Thus, what must be dealt with immediately must be dealt with as efficiently as possible before turning attention toward those things that will be dealt with later in the future. The presence of a denser cell population in the LVF may be one way in which the visual system has evolved to deal with immediate problems as efficiently as possible.

Perhaps the most compelling evidence for a functional division between the UVF and LVF comes from the neuroscience literature. Neurophysiological studies have determined that there is a direct neural link between the UVF and the ventral pathway of visual processing, largely involved in the processing of visual information for perception (Previc, 1990; Danckert & Goodale, 2001). These studies have also shown that there is a corresponding neural link between the LVF and the dorsal pathway of visual processing, largely involved in the processing of visual information for action. These neural connections are consistent with the specializations postulated for the UVF and LVF. The presence of the ventral and

dorsal pathways are central to the study of the two-visual systems hypothesis, which will be discussed in greater detail in the next chapter (Chapter 7). However, in the context of the UVF and LVF, the presence of direct and distinct connections from both visual fields suggests that there is a corresponding anatomical basis for the ecological arguments that postulate specialized differences across the vertical visual fields.

## 6.2 Hypotheses and Predictions

With respect to HCI, the functional specialization of the UVF and LVF suggests that user performance might be affected depending on where interactive elements on a display are initially perceived. If this is true, then the differences attributed to the UVF and LVF could be important in understanding how separate mental representations for perception and action relate to the organizational characteristics of a user interface and perhaps what kinds of interface layouts are optimal for specific user activities. Based on the existing psychological evidence, the following experimental hypotheses can be postulated:

1. User performance in pointing tasks should be measurably improved when the items to be selected are presented in the LVF compared to when equivalent items are presented in the UVF. This is consistent with the claim that although the UVF still enables motor movements such as pointing, the LVF is functionally specialized to facilitate these and other visually-guided motor responses.
2. There may be characteristic differences in the way that the LVF advantage exhibits itself between different kinds of pointing, for example direct (touch) pointing versus indirect (mouse) pointing. This is consistent both with previous experimental evidence that has consistently shown improved performance when physically pointing at objects in the LVF and also with previous comparisons of direct versus indirect input performance in HCI. Substantially less is known about how the UVF and LVF might contribute to differences in pointing performance with less direct input devices like mice, but it is worth noting that the visual fields became functionally specialized to deal with *physically direct* input but not necessarily input that depends upon an indirect mapping.

## 6.3 Experiment 2: Item Selection in the UVF and LVF

To test these predictions, a controlled experiment was designed to compare mouse and touchscreen pointing performance across the two visual fields. Performance was modeled using Fitts's Law (discussed in Section 3.1.2) as a means of characterizing movement times for pointing. If systematic performance differences existed depending on whether items were perceived in the UVF or LVF and whether items were pointed at with either mouse or touchscreen input, such differences would become apparent in the Fitts-regression models developed for each visual field and input device. The use of an experimental task amenable to Fitts's Law also meant that the experiment could be designed to test user performance in an interactive scenario consistent with previous research in HCI.

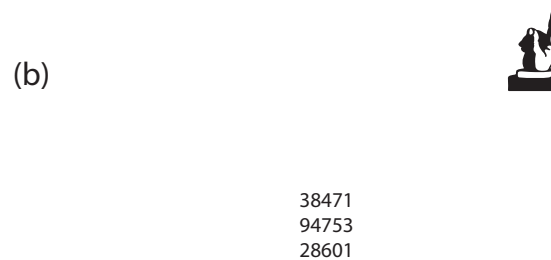
A fully counterbalanced, within-subjects experimental design was employed. Participants were instructed to point at individual items of different widths and at different distances as they were presented on a large-screen display. Every participant completed two blocks of pointing trials, with each block corresponding to either direct or indirect input. One block consisted of mouse pointing trials and another block consisted of touchscreen pointing trials. These blocks were counterbalanced such that half of the subjects completed mouse pointing before touchscreen pointing while the other half completed touchscreen pointing before mouse pointing. Figure 6.2 presents an illustration of a typical pointing trial in this experiment.

### 6.3.1 Participants

Eight participants were evaluated in this experiment. These participants were volunteers solicited from the undergraduate and graduate student populations at the University of British Columbia in a manner consistent with university ethical guidelines for participant selection. Five were male and three were female. Their ages ranged from 19 to 40 years. To minimize experimental bias due to handedness, all of the participants were required to be right-handed. All participants had normal or corrected-to-normal vision.

### 6.3.2 Apparatus

Figure 6.3 depicts the display and experimental apparatus used in this experiment. A SMART Board 3000i large-screen display was used to visually present trials to participants.



6

Figure 6.2: An illustration of a typical pointing trial in this experiment. (a) Participants were instructed to initiate trials by pointing at a square target on their right hand side. (b) A  $5 \times 3$  block of digits appeared either above or below eye-level. Participants would move their eyes to fixate on this block of digits and after a period of three seconds, this block of digits would drop out, be replaced by a single digit randomly chosen from 0 to 9, and a new square target would appear. (c) Participants had been instructed to make a univariate right-to-left pointing movement toward the center of this target while simultaneously saying the indicated number out loud. This “fixate-and-point” trial mechanism permitted control over whether selected items were perceived in the UVF or LVF without changing the biomechanics of the final pointing motion.





Figure 6.3: A photo of the experimental apparatus used in this experiment. Participants stood upright during each block of trials. In this photo, a participant is selecting items using touch input. In the mouse block of trials, a mouse was placed at a comfortable height on a stable surface to the right-hand side of participants.

The display was rear-projected and had an active LCD area of approximately 136 cm by 102 cm, running at a graphical resolution of 1024 by 768 pixels. A PC workstation (Pentium 3, 1.0 GHz machine with 256 MB of RAM running Microsoft Windows XP Professional Edition) drove the display, running the experimental software that presented trials and recorded participant performance data. The experimental software was written in C using Microsoft Visual C++, version 6. Graphical items were rendered using OpenGL and graphics hardware acceleration.

Movement times were captured using a high-resolution timer based on the CPU clock rate to achieve a sampling resolution of less than 1 ms. During the mouse pointing blocks, a standard Logitech optical Wheel Mouse was made available to participants. While participants were completing the experiment, a constant level of ambient (room) illumination was maintained and a researcher was present at all times.

Prior to each session, the SMART Board display was calibrated for touch input using the provided SMART Board driver software. All participants were instructed to stand in front of the display while the researcher manually adjusted a setting in the experiment software to ensure that presented targets to be selected in trials would always appear at eye level. This calibration for individual height differences ensured that every participant saw the same rendered display relative to their UVF and LVF, regardless of height.

### 6.3.3 Procedure

All participants took part in single sessions lasting approximately forty minutes in duration. In both mouse and touchscreen blocks, participants stood upright before the SMART Board at a viewing distance of approximately 30 cm (see Figure 6.3). Trials consisted of a single right-to-left pointing movement from a starting point to a single display target while fixating on a particular area of the screen whose vertical position was either above or below the displayed targets. These areas of fixations were used to experimentally control whether displayed targets appeared in the UVF or LVF.

Displayed targets were individual square targets that appeared in one of four different sizes:  $8 \times 8$ ,  $16 \times 16$ ,  $32 \times 32$ , and  $64 \times 64$  pixels. At the participant viewing distance of about 30 cm, these items subtended approximately 1, 2, 4, and 8 degrees of visual angle respectively. Targets appeared at one of four distances (or pointing amplitudes as they might be called in the literature related to Fitts's Law) from the starting position: 32, 64, 128, or 256 pixels. At the participant viewing distance, these yielded approximately 4, 8, 16, or 32 degrees of pointing movement respectively. Since all pointing movements were made from right-to-left along the horizontal dimension, the pointing task was effectively univariate, limited to one movement dimension at all times. Each combination of target size, target distance, and area of fixation was repeated, yielding a total of  $4 \times 4 \times 2 \times 3 = 96$  trials per block. Across both the mouse and touchscreen pointing blocks, every participant completed a total of 192 pointing trials.

All graphical items were rendered against a black background. Individual trials within blocks were initiated by instructing participants to point and hold their aim at a  $48 \times 48$  pixel starting square to their immediate right (see Figure 6.2). A  $5 \times 3$  array of randomly-generated digits between zero and nine appeared 135 pixels, or approximately 20 degrees of visual angle, above or below eye level. These arrays of numbers were rendered in a white,

---

fixed 10×10 font size. When participants fixated toward arrays located above eye level, presented items appeared in the LVF. Similarly, when participants fixated toward arrays located below eye level, presented items appeared in the UVF.

Participants were instructed to make pointing movements toward the center of displayed targets and were asked to emphasize speed and accuracy equally. They were also instructed to hold their gaze on the designated fixation area at all times in a trial, even while pointing. After a period of three seconds, the 5×3 array of numbers disappeared, replaced by just a single, randomly-selected digit between zero and nine in the same font and style simultaneously with the display of the target to be selected. Participants verbally indicated the single digit to the researcher while pointing at the displayed item. During the experimental design, it was decided that participants who incorrectly reported digits or were observed to shift their gaze while pointing would have those trials invalidated, although this never happened in the pool of participants who were evaluated in this experiment.

The vertical fixation and response mechanism served two purposes. First, it permitted experimental control of whether items appeared in the UVF or LVF. Second, it achieved this without changing the physical mechanics of the pointing interaction. This second point, in particular, was important to avoid any potential experimental confounds that might result from movements toward one direction or another. The chosen implementation of “fixate-and-point” provided sufficient assurance that participants maintained presented items in the UVF or LVF at all times. It might be argued that participants could “cheat” by fixating on the region of displayed target positions instead of the area of fixation. However, if they had done so, they would be unable to correctly report the final presented number in the fixation area while pointing because it would be very difficult to spatially individuate the number without re-fixating on it (Intriligator & Cavanagh, 2001). Participants would be at a considerable disadvantage if they employed other fixation “strategies” such as indicating the digit first and then re-fixating to point at the displayed target because this would slow them down considerably and would extend the duration of individual trials, thereby extending the overall length of the experiment. Participants were observed to be highly competent in completing trials and no participants were incapable of completing the experiment as instructed. Having participants verbally indicate numbers while pointing did not appear to interfere with the primary pointing task.

### **Mouse Condition**

In the block of mouse input trials, participants were provided with a Logitech optical Wheel Mouse. The mouse was placed on a stable surface that could be adjusted to comfortably accommodate participants of varying heights. Participants pointed at onscreen graphical items by aiming a rendered mouse cursor (the standard upper-left mouse pointer used in Microsoft Windows XP) at items and left-clicking them. During trials, the mouse cursor did not move from the starting position until the final displayed target appeared.

### **Touchscreen Condition**

In the block touchscreen input trials, participants directly pointed at onscreen graphical items by touching them with their index fingers. At the start of a trial, participants physically pointed and held their index finger down on the starting position until the final displayed item appeared. Participants pointed at the displayed item by lifting their fingers from the start position and tapping the item with the same finger.

### **Training**

All participants received a minimum of twenty practice trials prior to each block of experimental trials. These practice trials allowed participants to become familiar with the different pointing mechanisms for mouse and touchscreen input. Practice trials consisted of trials presented in the same manner as the experimental trials, but item sizes, item distances, and fixation location were randomized. Practice trials were presented until both participant and researcher were satisfied with a participant's ability to complete the task properly.

#### **6.3.4 Data Analysis**

Consistent with the literature related to Fitts's Law, least-squares linear regressions were used to analyze movement time against calculated indices of difficulty across aggregate pointing data from all participants. Separate linear regressions were conducted for each combination of visual field (UVF or LVF) and input method (mouse or touchscreen). If there is a general advantage toward movement for items presented in the LVF, such an advantage should be observed as consistently smaller  $y$ -intercepts or slopes in the LVF regression

models relative to the  $y$ -intercepts and slopes in the corresponding UVF regression models.

Trend analyses and curve estimation regressions were performed to find lines of best fit for each of visual field and input method as a secondary analysis. Individual performance differences were analyzed as a further exploratory measure by conducting a series of linear regressions across the visual fields and input methods for each individual participant. Thus, eight additional linear regressions were generated, one for each participant.

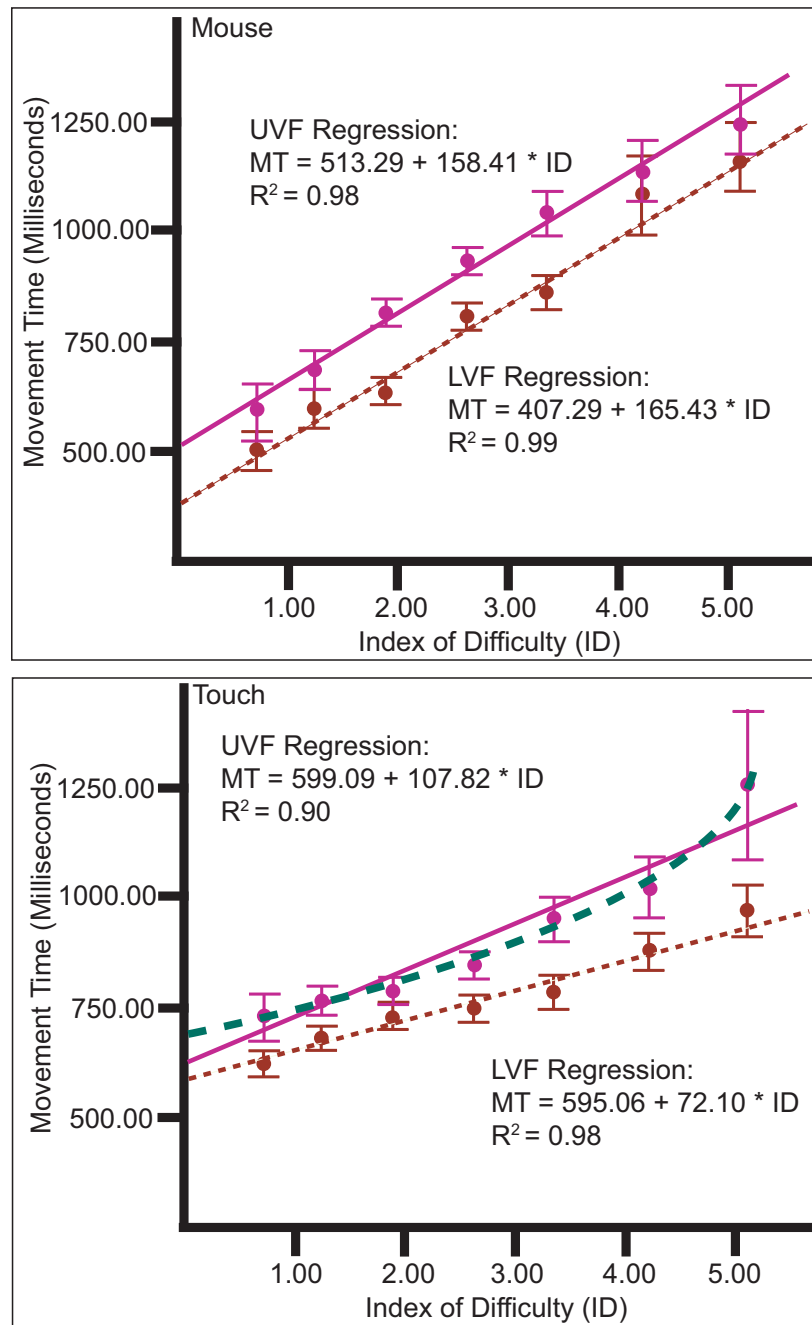
Pointing accuracy was quantitatively assessed using a root-mean-square error metric identical to the one used in Experiment 1 (see Chapter 5). This metric was applied to participant data and accuracy differences were analyzed using two-way ANOVAs with repeated measures (ANOVA-RMs) against visual field and index of difficulty for each input method. To characterize overall performance averaged across all combinations of item size and distance, group means between the visual fields and input methods were further analyzed with paired-sample  $t$ -tests.

### 6.3.5 Movement Time Performance

Figure 6.4 presents linear regression plots for the derived indices of difficulty (ID) versus movement time (MT). Consistent with the outlined experimental hypotheses, there are systematic differences in movement time performance that can be attributed to variations in whether pointing movements were made toward items initially perceived in the UVF or LVF. The regression lines indicated by the plots show that participants were consistently faster when items were presented in the LVF, even at very high indices of difficulty. The plots also demonstrate that this difference can be found regardless of input method.

Both mouse and touchscreen input were found to be consistently faster in favour of the LVF, though the effects were observed to be more pronounced with touchscreen input. The resulting  $y$ -intercept values were consistently smaller for both mouse and touchscreen regressions, but only touchscreen input was found to have a noticeably shallower slope. This particular result also appears to be consistent with the outlined experimental predictions: differences were observed in the way that the LVF advantage characterized itself across mouse and touchscreen input.

For mouse input, the regressions for both the UVF and LVF were highly consistent with the performance patterns predicted by Fitts's Law. Both regressions yielded very high  $r^2$  values ( $r^2 = .98$  for the UVF and  $r^2 = .99$  for the LVF), which meant that the lines of best



Error Bars Show Mean +/- 1.0 SE

Figure 6.4: Linear regression plots for movement time (MT) against index of difficulty (ID), broken down by visual field and input method. The LVF demonstrates a consistent advantage in all conditions, as shown by the more favourable LVF regressions. The linear fit generated for the UVF has a poorer fit for touchscreen pointing compared to all other regressions. A combined linear and quadratic curve (the dashed curve in the plot) fits best for this particular condition.

fit correlated very well with the sampled pointing data from participants. For touchscreen input, the LVF regression was also consistent with Fitts's Law, yielding a similarly high  $r^2$  value ( $r^2 = .98$ ). However, the UVF regression had a relatively poorer fit ( $r^2 = .90$ ). This suggested that a non-linear regression might provide a better fit to the sampled data and that Fitts's Law may not necessarily be the best performance model when comparing pointing performance across the visual fields.

A trend analysis across the visual fields and input methods provide statistical support for this observation. In the case of the UVF and touchscreen pointing, a trend analysis for polynomial contrasts indicated that a combined linear and quadratic fit would account for a greater proportion of the variance (contrast  $p < .001$ ; deviation  $p < .001$ ) than a linear fit could account for alone (linear  $\eta^2 = .90$ ; linear and quadratic  $\eta^2 = .98$ ). Contrasts of higher order than a quadratic fit were not statistically significant (all  $p > .05$  for higher order contrasts in the UVF touchscreen condition) and only linear contrasts were statistically significant for all of the other combinations of visual field and input method ( $p < .001$  for linear fits in all other conditions).

A subsequent curve estimation regression was performed to describe the polynomial coefficients for the combined quadratic and linear fit for the UVF touchscreen condition. This regression agreed with the trend analysis, indicating that a quadratic equation would fit better than a linear regression equation (quadratic equation:  $MT = 737.99 + 26.06 * ID^2 - 36.83 * ID$ ;  $r^2 = .98$ ). The resulting curve fit for this equation is shown in Figure 6.4. Thus, these data seem to indicate that Fitts's Law was obeyed with less fidelity under touchscreen interaction in the UVF, especially at the highest index of difficulty. When index of difficulty is broken down into target width (W) and target amplitude (A) and evaluated independently, it appears the better fit of a quadratic curve is largely the result of increased movement time during selection of targets with the smallest width of 8 pixels (refer to Figure B.2 in Appendix B for these data).

### 6.3.6 Precision and Accuracy

Figure 6.5 is a line graph plotting index of difficulty (ID) against RMS index. Because Fitts's Law is a model of movement time and does not make any predictions about pointing precision or accuracy, appropriate statistical techniques were applied to characterize pointing precision and accuracy across the visual fields and input methods. Nevertheless, one

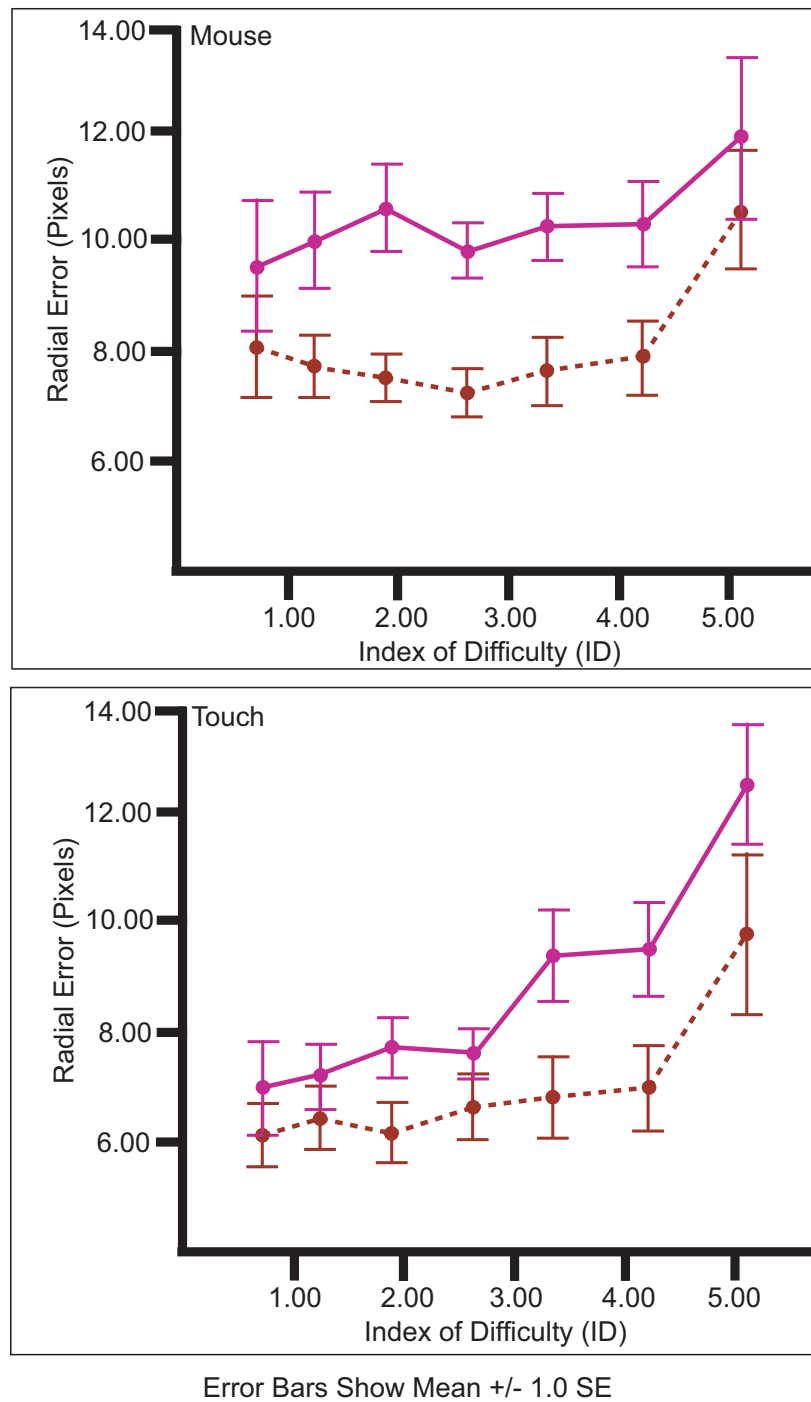


Figure 6.5: Line graph plots for pointing errors as measured by RMS index against index of difficulty (ID), broken down by visual field and input method. The UVF is represented by the solid line while the LVF is represented by the dotted line. Similar to the regressions for movement time, the LVF shows consistently better performance when compared against pointing to displayed items in the UVF.



would expect that there would be a general trend toward greater pointing errors as item size decreased and item distance increased (resulting in corresponding increases in the index of difficulty). The results from participant data were similar to those seen when comparing movement time performance across the UVF and LVF.

Pointing errors were consistently smaller when pointing to items perceived in the LVF. The two-way ANOVA-RM against visual field and index of difficulty for each input method indicated that there were statistically significant main effects for visual field [ $F(1, 21) = 15.294, p < .001$  for mouse input, and  $F(1, 21) = 10.204, p = .004$  for touchscreen input] and index of difficulty [ $F(6, 126) = 2.667, p = .018$  for mouse input, and  $F(6, 126) = 4.577, p < .001$  for touchscreen input]. Higher-order interactions between visual field and index of difficulty were not observed for either mouse or touchscreen input. Thus, these results suggest that displayed items with higher indices of difficulty were more difficult to acquire precisely. Moreover, the independent main effect of visual field indicates that perceived location relative to fixation was also important. It was possible to alter pointing precision by simply changing the visual field in which displayed items appeared.

### 6.3.7 Overall Performance Across Input Methods

Group means between the visual fields and input methods were analyzed with paired-sample  $t$ -tests to characterize overall movement time differences. These tests were statistically significant, indicating faster movement overall when participants pointed at displayed items in the LVF [ $t(7) = 3.488, p = .010$  for mouse input, and  $t(7) = 4.259, p = .004$  for touchscreen input]. Mouse pointing exhibited a mean movement time difference of 106 ms and touchscreen pointing exhibited a larger mean movement time difference of 143 ms between pointing in the UVF and LVF. These group mean differences were also observed with respect to the RMS measure of pointing error. Paired-sample  $t$ -tests in this regard were also statistically significant, indicating less error when participants pointed toward displayed items in the LVF [mouse  $t(7) = 2.500, p = .041$  and touch  $t(7) = 3.641, p = .008$ ]. The magnitudes of these differences were measured to be 3 pixels of radial error for mouse pointing and 2 pixels of radial error for touchscreen pointing between the UVF and LVF.

Figure 6.6 presents bar graphs showing overall performance across the visual fields, broken down by input method. These bar graphs illustrate that the systematic performance differences uncovered by the visual fields has a measurable influence on user performance,

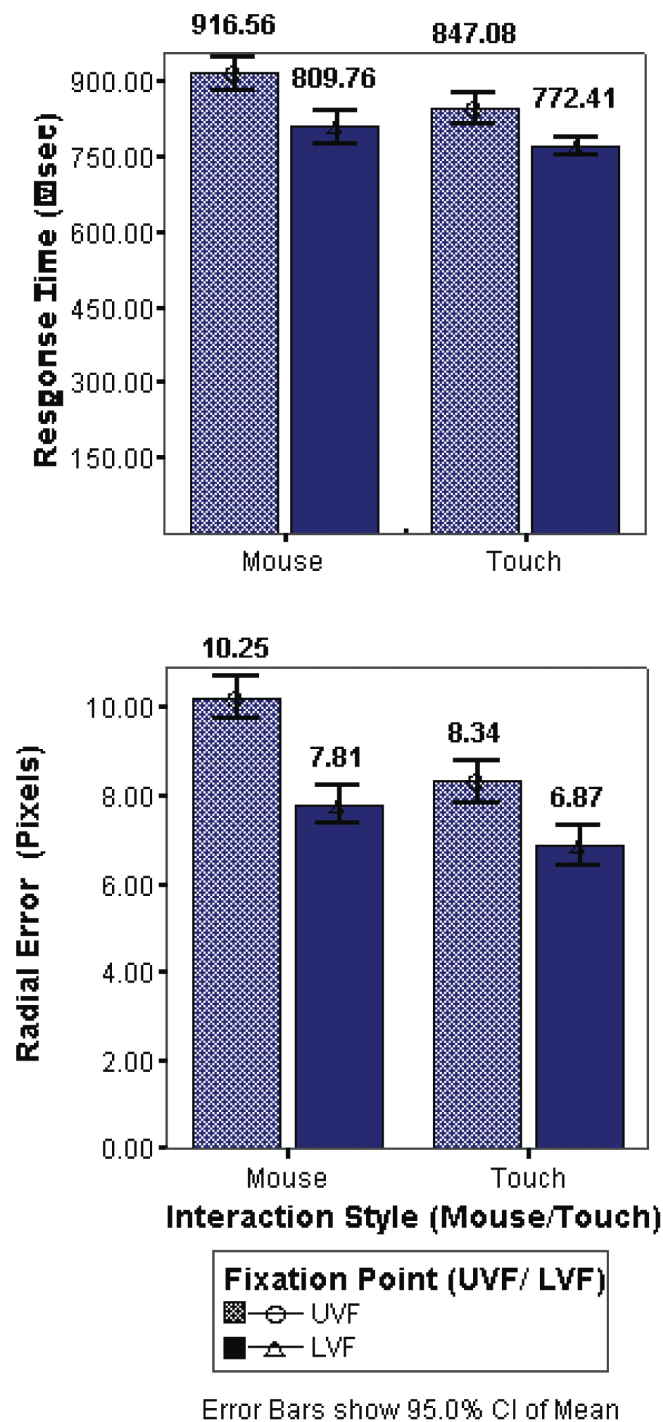


Figure 6.6: Bar graphs of overall movement time and pointing errors, broken down by visual field and input method. These bar graphs demonstrate that the influence of the visual field effects are apparent independent of the item size and distance, further suggesting their importance in user interface design.

---

independent of item size and item distance. These may suggest to interface designers and other practitioners that the influence of the visual fields cannot necessarily be ignored, since the cumulative impact of the visual field effects appears to be fairly substantial. In relative terms, pointing to items that were initially perceived to be in the LVF was 11.5 percent faster and approximately 29 percent more accurate than pointing to items initially perceived in the UVF when using a mouse. Pointing was 16 percent faster and approximately 24 percent more accurate when items were initially perceived in the LVF when pointing via touch.

### 6.3.8 Other Observations

The analysis of individual performance differences was consistent with the aggregate analyses described in the previous sections. The linear regressions from all eight participants yielded regression equations that were favourable to pointing performance in the LVF and this was true regardless of mouse or touchscreen input. Even so, there was considerable variation in the regression coefficients, although none were inconsistent with the outlined experimental hypotheses. Three of the eight participants were individually observed to have a trend toward non-linear increases in movement time in the UVF with touchscreen interaction, which may further account for the non-linear fit observed earlier. Although it might be easier to say that this particular finding is the result of statistical outliers, the observation that it occurred in nearly half of the participants suggests that this may instead be a symptom of characteristic individual differences between different participants.

Regardless of the specific quantitative differences, all eight participants agreed that they felt it was easier to point toward items when they were presented in the lower half of what they saw. Similar to the more qualitative findings from the previous experiment, none were able to give an adequate reason for this advantage, though it is adequately described and predicted by the representational theory. Moreover, several of the participants noted that these differences were not things they had noticed before, consistent with the intuitive view of a unified visual experience that is not necessarily shared by a representational approach to vision and HCI.

## 6.4 Discussion of Experiment 2

The experimental results clearly support a systematic manipulation of user performance based on separate representations of visual space as described by the functional differences in the UVF and LVF. The experimental hypothesis that displayed items are more efficiently selected when they are perceived in the LVF could have substantial implications for user interface design in existing and future systems. The non-linear fit observed in the UVF for touchscreen input and the characteristic differences between mouse and touchscreen input in the context of the visual fields also lend evidence to the experimental hypothesis that there are implicit differences in the way that visual information is processed when directly pointing versus indirectly pointing with a mouse. This further supports the claim that a representational approach to HCI will become more important as the design of many systems move toward more direct styles of interaction.

The continual shift toward ubiquitous and immersive computing environments makes the separation of visual space for perception and action a relevant design factor in HCI. The differences suggested by the present experiment on the UVF and LVF emphasize that the combined influence of attentional focus, eye position, item placement, and relative location can have a substantial impact on user performance. Further research into the UVF and LVF and how the differences between the two can be exploited may be useful in identifying design methods that implicitly reduce cognitive load in perceptually-complex user interface situations.

Furthermore, the functional dichotomy between the UVF and LVF suggests that there may be an important theoretical limit for direct visually-guided motor performance based on the present results. Although more research is required in this area, it seems plausible that more effort is required to interact when items are located in the extreme visual periphery, especially when these items are perceived in the UVF. In immersive or other large-screen environments, where graphical elements of particular importance may not be in the foveal view, understanding the differences between the UVF and LVF may provide insight into how graphical elements could be arranged to minimize the impact of unwanted effects.

In the future, a more thorough understanding of the differences between the UVF and LVF could lead to interesting implications for new technology intended to support user interaction. Future advances in eye-tracking technology might benefit from an understand-

---

ing of the differences between the UVF and LVF. Adaptive eye-tracking might be used to optimize the graphical rendering process for highly complex scenes in algorithmic strategies for perceptually-based rendering. Likewise, the UVF and LVF could be combined with predictive eye tracking and knowledge of gaze direction to dynamically optimize the layout of an interface for specific activities. This might be useful for maintaining a consistent sense of user flow, where interactions might conceivably be “pipelined” to improve performance. In these kinds of interfaces, current interactions might be implicitly executed because of their perceived location in the visual world while the user is already planning subsequent interactions for the future.

### 6.4.1 Implications for User Interface Design

The results from the current experiment suggest a number of strategies and implications for user interface design:

1. *The most frequently selected and most important interactive elements should be located in the lower half of the display.*

Items that are located in the lower half of a display are more likely to be initially perceived in the LVF by virtue of their position relative to the user. The experimental results suggests that following this simple placement strategy could help to optimize user performance by facilitating use of the LVF for interaction. Based on previous work related to the UVF and LVF, this might be especially pertinent for interface elements that demand greater attentional resolution where it is believed the LVF is functionally specialized.

2. *Physically direct interactions, such as touchscreen pointing, can be more efficient than less direct interactions, such as mouse pointing.*

The results from this experiment also provide further evidence for an advantage for physically direct interactions relative to less direct forms of input. The UVF/LVF literature in particular emphasizes that the visual fields have evolved to support physically direct input, as is found in interaction styles like touchscreen pointing, but not necessarily for less direct input. Following this particular guideline could help optimize user performance by facilitating use of this advantage. This is especially important for situations where users may need to select individual items rapidly and repeatedly.

3. *It is worth considering a strategy of organizing an interface for perception in the UVF and interaction in the LVF.*

Though the current experiment did not explicitly test any functional advantages in the UVF, the current results are consistent with other results from the UVF/LVF literature. Based on what has been studied here and elsewhere, designers might be able to take advantages of the differences between the visual fields to help develop an optimal organization of perceptual and interactive elements in a user interface design. As a hypothetical example, one could imagine an interface where all of the elements requiring the greatest saliency (i.e. warning lights, critical error messages) would be located in the upper half of a display while those requiring the greatest movement efficiency (i.e. toolbars, frequently used lists) would be located in the lower half of a display. This might be especially valuable when an interface is visually complex and conceivably demands a great deal of users' cognitive and attentional resources.

## Chapter 7

# The Two-Visual Systems

## Hypothesis <sup>1</sup>

An important thread in the study of visual processing is the proposal that visual information can be simultaneously represented in multiple ways. As aptly demonstrated by the functional specialization of the UVF and LVF (Chapter 6), it is possible to show that there are separate mental representations for visual perception and visually-guided motor action. In the case of the UVF and LVF, it would appear that one way to determine whether a visual task will be represented as perception or action is to identify where objects of interest are perceived in the visual world of an individual. However, the existing evidence in cognitive psychology suggests that this determination can be more than merely a function of spatial location. Starting with neuroanatomical evidence in the late 1960s, and a variety of experimental evidence in more recent years, a deeper model of differential processing of visual information for perception and action has emerged.

In this chapter, we will investigate the idea that mental representations of visual space are not only integral or separable, but that they also contain an important element of mediation as well. “Mediation,” in this sense, specifically refers to how a decision is made during visual processing: whether visual information should be treated as explicit perceptual information or as implicit information to enable appropriate physical movements. We will see how mediation can ultimately affect user performance under certain circumstances. Consistent with the empirical results presented in the previous chapter, the dichotomy between mediated perception and action can yield systematic differences in a user’s comprehension of visual space and subsequent user performance. Although this model occasionally goes by

---

<sup>1</sup>A version of this chapter has been accepted for publication. Po, B. A., Fisher, B. D., and Booth, K. S. (2005). A two visual systems approach to understanding voice and gestural interaction. *Virtual Reality*. In Press. A preliminary version of this experiment has also been published. Po, B. A., Fisher, B. D., and Booth, K. S. (2003). Pointing and visual feedback for spatial interaction in large-screen display environments. *Proceedings of the 3rd International Symposium on Smart Graphics*, pp. 22-38.

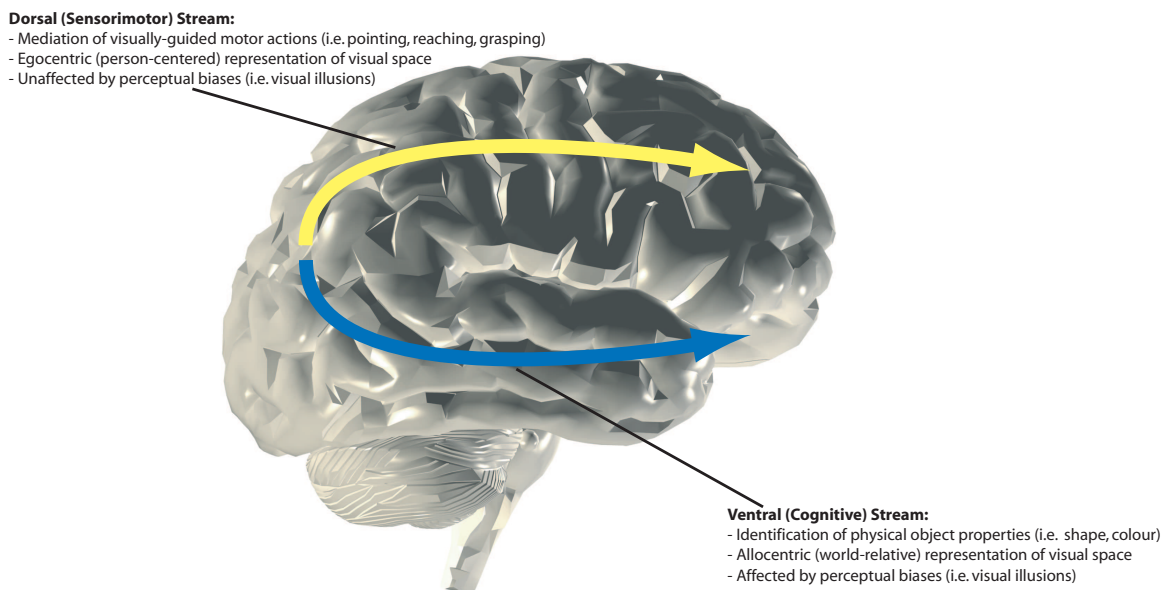


Figure 7.1: An illustration of the neuroanatomical division between the two-visual systems. The ventral and dorsal streams encapsulate two separate mental representations of visual space. The ventral stream (lower arrow) maintains the perceptual representation of space, necessary for activities like object identification while the dorsal stream (upper arrow) maintains the sensorimotor representation of space, necessary for motor activities like pointing, reaching, and grasping.

various names, including the “two-streams” or “dual-systems” theory of vision, we will refer to this model as the *two-visual systems hypothesis*, which refers to the postulated presence of two distinct neuroanatomical systems for visual processing.

## 7.1 Background and Related Work

The central argument of the two-visual systems hypothesis revolves around the postulated presence of two separate neuroanatomical systems, or *streams*, for processing visual information. From the perspective of a computational theory of mind, these two systems correspond to the “hardware implementation” of two distinct mental representations. One of these systems is known as the *ventral stream* of visual processing in the neuroscience literature. It is believed to process visual information in a manner consistent with a perceptual representation of visual space. The other system is known as the *dorsal stream* of visual processing. It is believed to process visual information in a manner consistent with an sensorimotor representation of visual space.



---

Figure 7.1 outlines the neuroanatomical division between the two visual representations. The ventral stream maintains an allocentric, or “world-relative”, representation of visually-perceived objects in the surrounding environment. The dorsal stream maintains an egocentric, or “body-relative,” representation of these same objects. Consistent with the evolutionary arguments outlined in the previous discussion of the UVF and LVF, the neuroanatomical separation of visual processing into these two distinct streams is believed to be the result of a biological need to use visual information to accomplish many different kinds of tasks.

Evidence suggests that the ventral stream is primarily responsible for enabling an *explicit*, accessible comprehension of the visual world required for visual tasks such as object identification and the parsing of complex visual scenes, which includes the perception of physical object properties such as colour and shape. This has led proponents of the two-visual systems hypothesis to suggest that an appropriate name for the ventral stream is a “what” system of visual processing. Likewise, evidence suggests that the dorsal stream is primarily responsible for enabling a rather *implicit* comprehension of the visual world required for enabling visually-guided movements, especially for physical actions that take place within peripersonal space (i.e. the visual space within reaching distance), such as pointing, reaching, and grasping. Consequently, the dorsal stream is often characterized as a “how” system of visual processing.

The definitions of the two, independent representations of visual space are so far the same as those used to define the representations of perception and action tied to the UVF and LVF of the previous chapter. However, unlike the functional differences observed across the visual fields, the two-visual systems hypothesis emphasizes that these differences can be the result of factors other than simply spatial location. Evidence in support of the hypothesis comes primarily from anatomical brain studies of humans and other primates, as well as controlled experimental work in cognitive psychology. Many of these studies have attempted to separate those visual tasks that are driven by the ventral or perceptual representation of visual space from those that are driven by the dorsal or sensorimotor representation of visual space. It is this special aspect of visual mediation that is of interest to HCI.

Although the egocentric representation provided by the dorsal stream is specialized for visually-guided movement tasks, experimental evidence suggests that either the ventral

or dorsal stream can provide the information necessary to make skilled motor movements (Bridgeman et al., 1997). One of the most recent reviews of the two-visual systems hypothesis by Goodale and Milner (2004) suggests that the ventral stream encapsulates an enduring representation of objects and their spatial relationships while the dorsal stream encapsulates a moment-to-moment representation appropriate for rapid visual response. The determination of what representation is most influential in configuring the final motor response is dependent on a variety of factors, including response delay and the presence or absence of particular visual cues.

Understanding the different kinds of visual characteristics that “trigger” one representation over another, some of which have already been identified in the psychological domain, could be particularly valuable in the study of HCI as they might help practitioners understand what kinds of user interaction are inherently driven by perceptual or sensorimotor representations of visual space. In the sections that follow, we will investigate how existing forms of user input can move from being driven by one representation to another by simply changing a single individual characteristic of a visual display. The interplay between separate mental representations of visual space offers the opportunity to predict user performance patterns in a way that might not otherwise be possible, including the characterization of particular kinds of user errors that designers may wish to help users avoid.

### 7.1.1 Evidence from Neuroscience

The majority of the evidence supporting the two-visual systems hypothesis is derived from studies of split-brain monkeys by Trevarthen (1968). The results of his experimentation led him to believe that the perception of object location and object identity were subserved by anatomically distinct brain mechanisms. He originally called these *focal* and *ambient* systems, with the focal system referring to what is known here as the ventral stream while the ambient system referring to what is known here as the dorsal stream. Trevarthen’s terminology was replaced by less ambiguous terms in subsequent years because the terms focal and ambient became better known to refer to foveal and peripheral vision respectively.

Schneider (1969) independently came to the same conclusions as Trevarthen through work with brain-lesioned hamsters. Their collective work provided the basis for much of the neuroscientific research on the neural mechanisms underlying perception and action for the better part of the 1970s and 1980s. Subsequent research by Ungerleider and Mishkin

---

(1982) brought into common usage the ventral and dorsal terminology, and they are credited with originally defining these streams of visual processing as “what” and “where” streams, respectively. Goodale and Milner (1992) reviewed evidence from case studies with brain damaged patients as well as the results of other work to further characterize the ventral and dorsal distinction as being instead “what” and “how” streams.

Perhaps some of the most convincing evidence in favour of the two-visual systems hypothesis are the direct case studies with patients who exhibit symptoms of *optic ataxia* (a severe deficit in making appropriate movements toward objects although patients have no trouble in identifying or classifying them), *visual agnosia* (an inability to properly identify objects although patients have no difficulty in making movements toward them), and *blindsight* (a severe visual disorder characterized by an inability to “see” while retaining the ability to make appropriate movements). Jeannerod (1986), and Perenin and Vighetto (1988) are credited with reporting some of the classic examples of optic ataxia. More recent evidence by Pisella and Rosetti (2000) suggests that a disruption of online motor control partially localized to the dorsal pathway of visual processing is responsible for optic ataxia, consistent with a two-visual systems model of perception and action.

Visual agnosia is often reported in case studies of patients who have had brain damage that is at least partially localized to the ventral stream of visual processing. Goodale and Milner (1992) report on patient DF, who is arguably the most “famous” subject of all those reported in case studies. Patient DF has been reported to be unable to recognize object size, shape, or orientation, although her prehensile motor skill has apparently remained unaffected. When instructed to pick up objects that she could not verbally identify, her ability to pick up these objects has been observed to be approximately equal in performance to those of a healthy individual. These remarkable results are similarly seen in patients exhibiting blindsight. Reports by Weiskrantz (1986), Perenin and Rosetti (1996), and Jackson (2000) have demonstrated that patients with blindsight can still reach out and orient their eyes and hands toward supposedly “unseen” objects in the surrounding visual world. All of these case studies strongly suggest that the neuroanatomical mechanisms supporting perceptual tasks are at least partially distinct from those mechanisms supporting visually-guided movement tasks, and that it is possible to identify differential performance depending on the visual task itself.

### 7.1.2 Evidence from Experimental Work

Concurrent with the research arising from the field of neuroscience, behavioural evidence consistent with a two-visual systems model also arose in the field of cognitive psychology. Numerous psychophysical experiments have been conducted since the late 1970s, further demonstrating that it is possible to segregate visual space representations of perception from visual space representations of action. These experiments are in some ways more important to HCI than the evidence from neuroscience, because they demonstrate that even healthy individuals are capable of differential visual task performance consistent with the two-visual systems hypothesis, and that it is possible to devise situations where one can systematically shift individuals from using one mental representation of space to another simply by altering the fundamental characteristics of the task at hand.

One of the earliest two-visual systems experiments was a saccadic eye movement study by Bridgeman, Hendry, and Stark (1975). They found that subjects could not reliably report a change in target position if the change was timed to take place in the middle of a visual saccade. Nevertheless, these subjects were always able to *point* at the positions of these targets, regardless of whether they were able to verbally report a displacement. A subsequent experiment by Bridgeman et al. (1979) demonstrated that the same tasks could be “assigned” to either the ventral or dorsal visual processing streams, suggesting that a distinction between the two systems could be observed by simply looking at the responses that each system was postulated to provide.

These two experiments gave rise to a series of experiments exploring rapid “online” arm movement and control. A possible confound in the earlier studies was the presence of a visible hand, which could have affected the experimental outcome. Pelisson et al. (1986) found that the results of earlier experiments could be replicated, even with the additional constraint that subjects were unable to see their own hands. In a target displacement experiment, they found that subjects were still capable of altering their movements to compensate for verbally unreported shifts in target position. Furthermore, they found that subjects consistently reported that they were unaware of making such movement changes. These results have been influential in subsequent experiments involving the two-visual systems hypothesis.

---

## 7.2 Visual Illusions

An important component of most of the experimental work related to the two-visual systems hypothesis in cognitive psychology is the use of *visual illusions*, or apparent ambiguities in the visual world, to separate the visual representations of perception and action. The evidence from these psychological experiments suggests that the world-relative view provided by the ventral stream is susceptible to difficulties when dealing with egocentric judgments and that visual illusions can affect a person's ability to make such judgments. This is often used to explain why visual illusions only appear to affect ventral stream responses. Because the dorsal stream is characteristically egocentric, it has no difficulty making such judgments. This distinction in behaviour suggests that knowing what kinds of visual features influence the ventral and dorsal representations of space could be helpful in predicting performance differences in certain kinds of task scenarios, especially where a choice between perceptually-based interactions like voice input and movement-based interactions like pointing must be made.

Bridgeman, Kirch, and Sperling (1981) used the visual phenomenon of induced motion to extend earlier findings about saccadic eye movements. In this experiment, a large structured background was continuously displaced during the visual presentation of a small target. They discovered that motor responses toward the presented target were substantially less affected by the apparent "movement" of the target than cognitive (i.e. verbal) responses about the target. Their results were interpreted as showing that induced motion only affected conscious perception while real target displacement only affected motor behaviour. Aglioti, DeSouza, and Goodale (1995) demonstrated that the size-contrast ambiguities found with Ebbinghaus-Tischner circles were reliably reported by subjects when asked for a verbal indication of size, but when subjects were then asked to reach for the circles, their grip apertures remained constant and appeared to be largely determined by the true sizes instead of the apparent sizes. Other studies, such as those by Haffenden and Goodale (1998) and Gentilucci et al. (1996), report similar results with other kinds of visual phenomena.

### 7.2.1 The Induced Roelofs Effect

Several of the most recent experiments involving the two-visual systems hypothesis in cognitive psychology have involved the use of a visual illusion known as the *induced Roelofs Effect*, which has visual elements similar to the ones that are seen in many kinds of graphical display settings, such as large screen graphical user interfaces and immersive virtual environments. The “induced” Roelofs Effect is derived from a more general phenomenon first identified by Roelofs (1935). It is perhaps best described as the systematic misperception of target objects that are presented within the decontextualized visual field of an individual. Instances of such an effect have been observed in various circumstances. In particular, Mateef and Gourevich (1983) observed a tendency to perceive the locations of flashing lights as being closer to the line of sight than their actual positions. Bridgeman, Peery, and Anand (1997) have used the induced Roelofs Effect to study performance differences between ventral and dorsal stream responses. In their study, they found that certain kinds of report, such as verbal response, were affected by the induced Roelofs Effect while motor forms of report, such as distal pointing, remained unaffected.

The use of the word “induced” is meant to make a distinction between the evoked illusions and the underlying principle behind their perceptual effects. Figure 7.2 illustrates the perceptual effect of the induced Roelofs Effect used in experiments involving the two-visual systems hypothesis. This particular illusion can be described as a systematic bias in the perceived location of objects that have been presented within a surrounding rectangular frame. When a rectangular frame is asymmetrically displaced by some offset distance to the left or right of a viewer, a perceived bias in the location of the objects within the rectangular frame occurs. When the frame is offset to the left, objects within the frame are systematically perceived as being further to the right of the viewer than they really are. Likewise, when the frame is offset to the right, objects within the frame are systematically perceived as being further to the left than they really are. When the frame is centered, there is no perceptual bias and the presented objects are consistently perceived in their correct locations.

Understanding visual illusions like the induced Roelofs Effect may be important in the context of user interface design. The rectangular frame of the illusion is similar to bordering elements, such as virtual window frames or the physical walls of a large screen

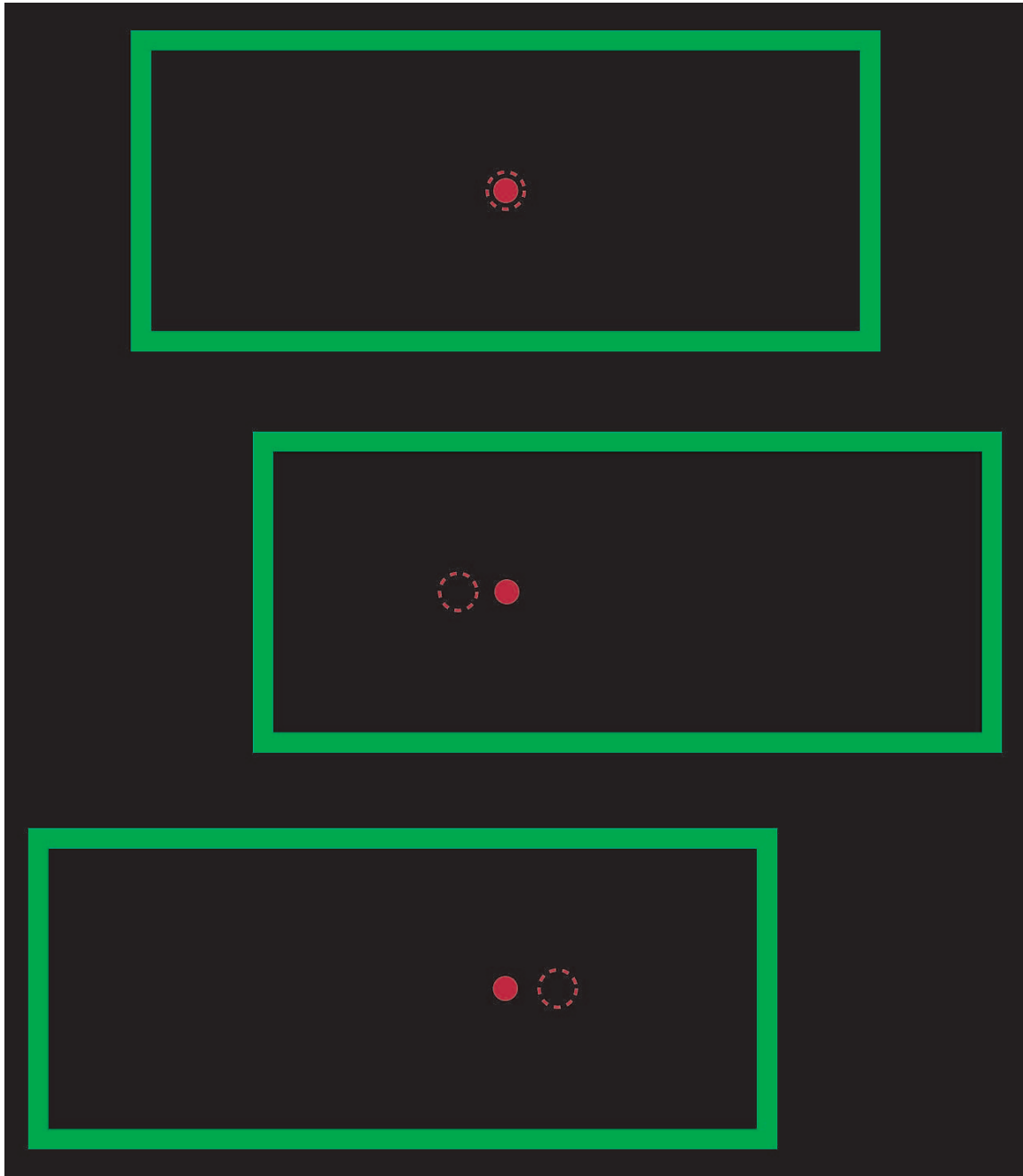


Figure 7.2: An illustration of the induced Roelofs Effect. When objects are surrounded by an offset rectangular frame, they appear more to the left or right of center than they really are. In the figure, solid circles represent actual target positions while dashed circles represent perceived locations.

display, which provide visual context in a graphical display. The objects within the frame are likewise similar to the icons and interactive elements of a GUI-based application, such as buttons, menu items, or other targets. In the experiment that follows, we will use the induced Roelofs Effect to study the influence of asymmetric frames in an immersive large screen environment and to understand how the addition or subtraction of certain kinds of visual cues can systematically move individuals from the use of one representation of visual space to another.

### 7.3 Hypotheses and Predictions

The “classic” predictions associated with the two-visual systems hypothesis suggest that responses drawn from a perceptual representation of space are susceptible to the biases of visual illusions while responses drawn from a motor representation of space are robust against these effects. Because the induced Roelofs Effect has been used in previous experiments of the two-visual systems, and because it contains basic visual elements closely related to the ones seen in GUI-based systems, this visual illusion is central to the hypotheses and predictions made here. Based on previous experimental evidence, the following hypotheses have been postulated:

- *Voice-based input* is a form of interaction that draws upon a perceptual representation of space and solely depends upon the ventral stream of visual processing because no direct, physical movement of the limbs is required for the response. Thus, this kind of interaction will be most susceptible to the perceptual ambiguities of the induced Roelofs Effect.
- *Pointing without visual feedback* (i.e. without any visible graphical cursor) is a form of interaction that draws upon a motor representation of space and solely depends upon the dorsal stream of visual processing because a direct, physical movement is required for response and there is no reliable way to make visual corrections to initial pointing movements. Thus, this kind of interaction will be unaffected by the perceptual ambiguities of the induced Roelofs Effect.

These two predictions are central to most of the experimental work on the two visual systems in cognitive psychology. Demonstrating that this kind of dissociation exists allows



---

researchers to infer the existence of two separate mental representations of visual space. In this experiment, these two basic predictions were extended to study how the two-visual systems hypothesis might extend to other kinds of user interaction. Two other experimental hypotheses were formulated:

- *Pointing with visual feedback* (i.e. with a visible graphical visual cursor) engages the ventral stream of visual processing, thereby making it at least somewhat dependent upon a perceptual representation of visual space. The presence of visual feedback means that participants will have an explicit visual awareness of their pointing movements not seen when pointing without visual feedback. Thus, the presence of visual feedback means that such “closed loop” interactions will be susceptible to the perceptual ambiguities of the induced Roelofs Effect.
- *Pointing with lagged visual feedback* (i.e. with a temporally-delayed graphical cursor) could engage either the ventral or dorsal streams of visual processing, depending on the interaction strategy employed by the user. Users who disregard the visual feedback will effectively make the pointing interaction an “open loop” interaction, like pointing without visual feedback while users who continue to depend on the visual feedback effectively make the pointing interaction a “closed loop” interaction, like pointing with visual feedback. Thus, the presence of lagged visual feedback could cause some participants to be susceptible to the perceptual ambiguities of the induced Roelofs Effect, while others might not be affected.

In the absence of a two-visual systems model of perception and action, these experimental hypotheses might seem counterintuitive. They predict that voice-based interaction will be more susceptible to perceptual errors than other kinds of physical interaction and they predict that pointing performance will be poorer in some way *with* visual feedback than when it is absent. Moreover, these hypotheses predict that a lag in displaying the visual cursor might actually *improve* performance compared to a non-lagged visual cursor. However, if the performance predictions made by the two-visual systems hypothesis hold in a setting typical of large screen interaction, they challenge some of the most basic design assumptions made when designing an interface. Most applications assume that reliable feedback in the form of a visual cursor is necessary for pointing and emerging multimodal techniques assume that voice-based input is not susceptible to errors induced by the pres-

ence of graphical frames. Furthermore, it is almost universally assumed that the presence of interactive lag is always detrimental to user performance.

## 7.4 Experiment 3: Visual Feedback on Large Screens

A controlled experiment taking place in a large screen, immersive environment was designed to test the outline experimental hypotheses. A simple target acquisition task was designed, where vocal localization and spatial pointing were equally feasible methods of interaction. In this experiment, participants were instructed to complete four blocks of trials requiring them to select presented targets from fixed positions. Each block of trials used a different mode of input for target selection. One block used voice-based input and three blocks used pointing under varying levels of visual feedback.

The experiment employed a within-subjects experimental design. Each participant attended a single, individual session lasting approximately one hour where all four blocks of trials were completed. Each block could be considered a distinct experimental condition, characterized by the use of a specific kind of interaction. Every block consisted of 48 trials, and every participant completed a total of  $4 \text{ blocks} \times 48 \text{ trials} = 192 \text{ trials}$ . Order of block presentation was fully counterbalanced such that each participant had a unique presentation order. Figure 7.3 illustrates the progression of a typical trial in this experiment.

The four experimental blocks or conditions were identified and characterized as follows:

1. *Voice-based input.* An effectively continuous voice protocol was used for target selection. No physical pointing interactions occurred in this experimental condition.

The three remaining conditions used a continuous, spatial pointing interaction for target selection. A handheld pointer (a Polhemus Fastrak simulating a ray-casting pointer, as in Experiment 1 of Chapter 5) was used. These conditions differed from one another by the kind of visual feedback provided for pointing during trials.

2. *Pointing without visual feedback.* No tracking cursor was visible during this experimental condition, meaning that participants were effectively “blind” to their pointing movements during this block of trials.
3. *Pointing with visual feedback.* A tracking cursor was visible during trial pointing. The

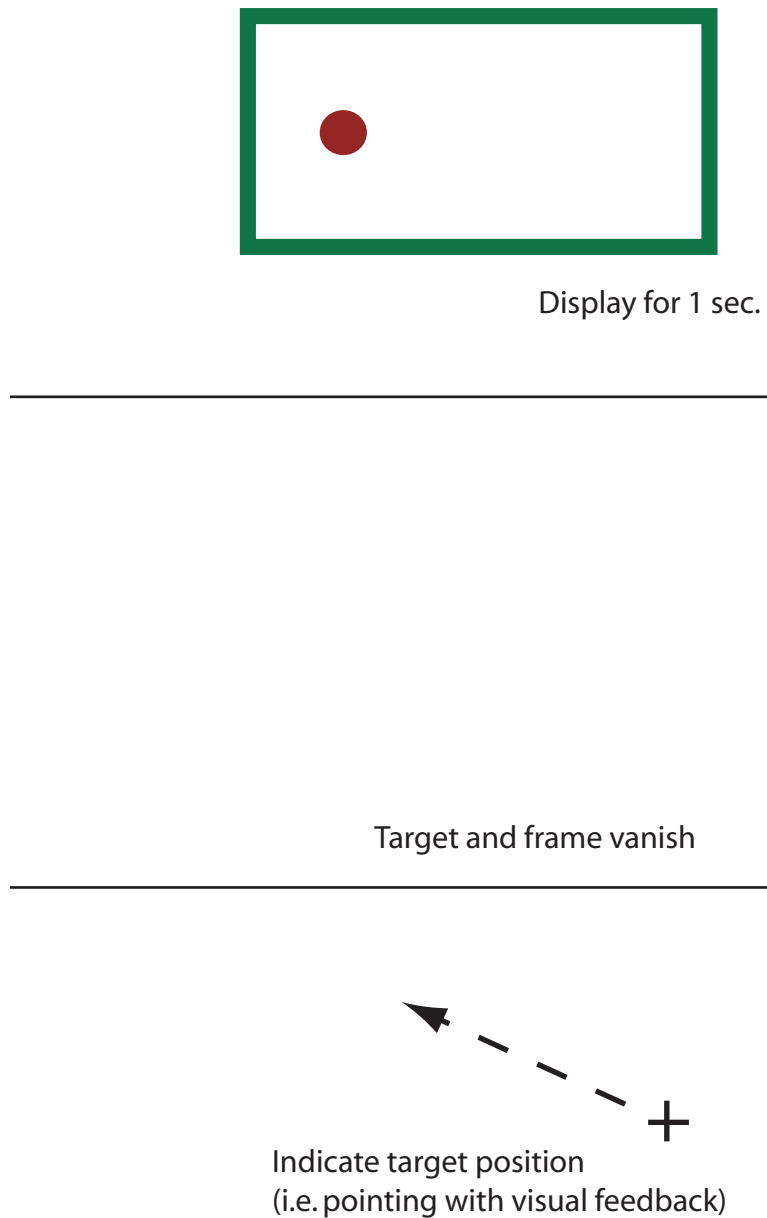


Figure 7.3: An illustration of a typical trial in this experiment. At the beginning of each trial, a red target would appear in a horizontal position, eight degrees of visual angle across the midline of a participant, surrounded by a green rectangular frame that was either centered on the participant, or offset to the left or right by four degrees of visual angle. After one second, the target and frame vanished, and participants were instructed to indicate the position of the previously presented target. The method of response varied by trial block.

cursor was a graphical crosshair similar to the kind of visual feedback often used in interactive desktop and virtual reality-style environments.

4. *Pointing with lagged visual feedback.* The tracking cursor used in the previous pointing with visual feedback condition was temporally-delayed during pointing. A one-half second lag was added to the cursor to simulate apparent latency. The intention of this condition was to identify the potential influence of lag on the perceptual and motor representations of space rather than to simulate the response lag typically seen during user interaction.

Consistent with the previous experiments described in previous chapters, the visual stimuli used in these experiments are generally more basic than those that would be found in “real-world” settings. Nevertheless, the choice of these stimuli was intentional because control over the visual characteristics of the display was necessary in a way that could not practically be achieved with a standard window-based GUI. Despite their basic nature, these stimuli conferred other benefits over more visually complex graphical displays. First, the simpler visual stimuli removed many of the possible confounding display factors that could be used to explain the performance differences between the four interaction techniques. Second, the use of this kind of display permitted the employment of a task that was representative of an entire class of user interaction. Third, using such a visual display offered an opportunity to understand how even the most basic visual elements can impact user performance.

#### **7.4.1 Participants**

Twenty-four participants took part in this experiment. Sixteen were male and eight were female. Their ages ranged from 18 to 31 years. All of the participants identified themselves as right-handed and as having normal or corrected-to-normal vision via self-report.

#### **7.4.2 Apparatus**

Figure 7.4 presents a photograph of the large screen display setting used in this study. Even though the display was a three-screen, wide-angle projection surface, only the center surface was used in this experiment. The active display was forward-projected and it had physical dimensions of 275 cm by 215 cm. Participants were seated at a distance of 250 cm to avoid



Figure 7.4: A photograph of the large screen display environment used in this experiment. Participants were centered and seated before a three-screen, wide-angle display. Only the center display was used in this study. Pointing interaction was implemented with a Polhemus Fastrak and an attached stylus. Arms and hands were kept beneath a large wooden table at all times. During sessions, all ambient light was extinguished and a researcher was always present.

projector occlusion effects. With the exception of illumination from the projector, all other remaining light sources were extinguished. During the experiment, a researcher was always present to facilitate session progression. A PC workstation (Pentium 3, 800 MHz desktop with an nVidia GeForce 3 card and 1 GB of RAM) and custom software written in Java 1.3 were used to render trials and record quantitative data. The projector display had an effective resolution of 1024 by 768 pixels.

A large, wooden table was constructed and positioned directly in front of participants to obscure viewing of their hands and arms during the experiment. The table had dimensions of 120 cm by 95 cm by 80 cm (width, depth, and height respectively). These table dimensions ensured that the participants would have enough space to make free distal pointing movements without obscuring their ability to see the display. This particular requirement of keeping their hands beneath the table throughout the experiment was intended to strictly control their perception of visual information to only that provided by the display although participants still retained access to proprioceptive information about their physical pointing movements. This requirement is consistent with previous experiments involving the

two-visual systems in cognitive psychology (Pelisson et al., 1986).

A Polhemus Fastrak was used to implement pointing in this experiment. Prior to each session, the Fastrak was calibrated and registered via the experiment software. All sources of metallic interference were kept away from the transmitter and sensors. A Fastrak freehand pointer (the same as used in Experiment 1) was held in the right hand of participants while a standard push-button mouse was held in the left hand of participants. The two-handed pointer and mouse setup was used to allow participants to point at the display with one hand while pressing a mouse button in their other hand to confirm their final pointing position. As in Experiment 1, this was used to enable the cleaner collection of experimental data than might be achieved through one-handed input or other alternatives, such as pointing and dwelling at an intended target. With only the attached pointer, the Polhemus Fastrak maintained an effective update rate of 120 Hz.

### 7.4.3 Procedure

Experimental trials consisted of a one-second presentation of a single, red circular target surrounded by a green rectangular frame on a black background (see Figures 7.2 and 7.3 for illustrations). The circular targets were one degree of visual angle in diameter and could appear anywhere along an eight-degree horizontal continuum centered on the participant (four degrees to the left, and four degrees to the right). The presented target position for a given trial was randomly selected from a uniform distribution such that no particular portion of the eight-degree continuum had more presentations than any other. The rectangular frames were 21.0 degrees in width by 9.0 degrees in height, with a line thickness of one degree. For any given trial, a presented rectangular frame was either centered on the screen relative to a participant's midline, or it was offset to the left or right of center by 4.0 degrees.

After one second, the target and frame vanished, leaving only the black background. Participants were instructed to respond immediately after the target and frame vanished by indicating the position of the now-extinguished target using the input method specified by the block of trials they were completing. This "extinguish and point" design was used to enable a controlled assessment of performance across all of the tested methods of interaction. For example, if the target remained present on the display, the pointing without visual feedback condition could not be fairly compared to the other interaction conditions and we would not be able to characterize any measured performance differences across perceptual

and motor representations of visual space with any certainty.

These trial parameters resulted in sixteen repetitions for each of the three frame positions, yielding the total of 48 trials per experimental block. Trial randomization was performed in a way such that no two consecutive trials in a condition had the same target position and frame position.

### **Voice Input Condition**

In the voice-based input condition, participants specified their judged position of presented targets on a nine-point scale. Voice “input” was simulated using a Wizard-of-Oz technique: participants called out their responses and the researcher who was present manually entered in the participant responses. The verbal indication “one” meant a judgment that was furthest to the left and the verbal indication “nine” meant a judgment that was furthest to the right. Participants were told to use whole numbers in their responses and that fractional values would not be accepted, even though targets could be at non-integral positions.

### **Pointing Conditions**

Pointing interactions were accomplished with the Polhemus Fastrak, attached pointer and push-button mouse. Responses were made by aiming the pointer with their right hand at the display like a laser pointer. Once participants were satisfied with where they were aiming, they pressed a button on the mouse held in their left hand to indicate target selection.

### **Training**

Participants were provided with instructions at the beginning of the session and prior to the start of each experimental condition. Each block of 48 trials was preceded by a minimum of ten practice trials where participants were provided with a chance to familiarize themselves with the response protocol for that particular block of trials. Practice trials were presented in the same manner as actual experimental trials, except that the rectangular frame remained fixed in a centered position. This was meant to prevent participants from gaining any additional “experience” or exposure to the induced Roelofs Effect before each condition.

#### 7.4.4 Data Analysis

To test the theoretical predictions outlined earlier, the experiment employed here made it relatively straightforward to conduct a quantitative analysis of participant performance. Determining whether a particular experimental hypothesis demonstrated evidence of validity could be rephrased as simply asking whether a given method of input exhibited the presence or absence of the induced Roelofs Effect in participants. Because it was predicted that voice-based input and pointing with visual feedback were primarily influenced by the ventral representation of visual space, participants were expected to be more “susceptible” to the perceptual bias of the induced Roelofs Effect. Similarly, because pointing without visual feedback and pointing with lagged visual feedback were predicted to be more influenced by the dorsal representation of visual space, participants were expected to be less “susceptible” to the induced Roelofs Effect.

Participant responses in each of the four conditions were defined as the horizontal ( $x$ -coordinate) offset to the left or right of center, measured in degrees of visual angle. A global one-way ANOVA was performed on each of the four interaction conditions to analyze responses to manipulations of frame position across all subjects. If the induced Roelofs Effect were present with the responses provided by a given interaction technique, this would show up as a statistically significant main effect of frame position. This analysis was complimented by individual one-way ANOVAs for *each* subject in each of the four interaction conditions, consistent with previous experiments involving the two-visual systems hypothesis in cognitive psychology and other techniques for measuring performance differences in psychophysical experiments (Bridgeman et al., 1981, 1997; Vicente & Torenvliet, 2000). This second analysis was specifically performed to characterize the prevalence of the induced Roelofs Effect in each of the four experimental conditions. Based on the experimental hypotheses, one would expect more participants to be affected by the induced Roelofs Effect in a statistically significant manner when responding with voice-based input and pointing with visual feedback.

Table 7.1 and Figure 7.5 present a summary of the results collected in this experiment. The table presents the results of each participant across all four types of tested interaction. The cells marked in bold indicate that a participant had a statistically significant main effect of frame position with a given interaction technique as assessed by the one-way ANOVA (i.e.



Participant	Voice-Based Input	Pointing w/ Feedback	Pointing w/ Lagged Feedback	Pointing w/o Feedback
1	<b>YES (p &lt; .001)</b>	<b>YES (p &lt; .001)</b>	NO (p = .951)	NO (p = .707)
2	<b>YES (p &lt; .001)</b>	NO (p = .484)	NO (p = .843)	NO (p = .985)
3	<b>YES (p &lt; .001)</b>	<b>YES (p &lt; .001)</b>	NO (p = .952)	NO (p = .975)
4	<b>YES (p &lt; .001)</b>	NO (p = .955)	NO (p = .697)	NO (p = .950)
5	NO (p = .765)	NO (p = .636)	NO (p = .988)	NO (p = .968)
6	<b>YES (p &lt; .001)</b>	<b>YES (p &lt; .034)</b>	NO (p = .752)	NO (p = .572)
7	NO (p = .537)	NO (p = .093)	NO (p = .939)	NO (p = .800)
8	<b>YES (p &lt; .001)</b>	NO (p = .789)	NO (p = .846)	NO (p = .963)
9	<b>YES (p &lt; .001)</b>	<b>YES (p = .012)</b>	NO (p = .358)	NO (p = .933)
10	<b>YES (p &lt; .001)</b>	<b>YES (p = .027)</b>	NO (p = .834)	NO (p = .428)
11	NO (p = .285)	<b>YES (p = .006)</b>	NO (p = .675)	NO (p = .932)
12	NO (p = .899)	NO (p = .856)	NO (p = .940)	NO (p = .964)
13	<b>YES (p = .049)</b>	<b>YES (p = .043)</b>	NO (p = .971)	NO (p = .954)
14	NO (p = .646)	NO (p = .539)	NO (p = .914)	NO (p = .804)
15	<b>YES (p &lt; .001)</b>	<b>YES (p = .036)</b>	NO (p = .642)	NO (p = .905)
16	<b>YES (p = .010)</b>	<b>YES (p = .018)</b>	NO (p = .993)	NO (p = .819)
17	<b>YES (p &lt; .001)</b>	NO (p = .792)	NO (p = .867)	NO (p = .999)
18	NO (p = .171)	NO (p = .978)	NO (p = .824)	NO (p = .790)
19	<b>YES (p &lt; .001)</b>	<b>YES (p = .015)</b>	NO (p = .791)	NO (p = .942)
20	NO (p = .318)	NO (p = .177)	NO (p = .179)	NO (p = .832)
21	<b>YES (p &lt; .001)</b>	<b>YES (p = .042)</b>	NO (p = .995)	NO (p = .998)
22	<b>YES (p &lt; .001)</b>	<b>YES (p = .004)</b>	NO (p = .946)	NO (p = .978)
23	<b>YES (p &lt; .001)</b>	<b>YES (p = .038)</b>	NO (p = .794)	NO (p = .531)
24	NO (p = .326)	<b>YES (p = .013)</b>	NO (p = .346)	NO (p = .817)
Total	16	14	0	0

Table 7.1: A table of individual participant performance in this experiment. This table provides an overview of the individual one-way ANOVAs for main effects of frame position. Bold cells indicate that a statistically significant main effect of frame position was found for the given participant with the given method of interaction.

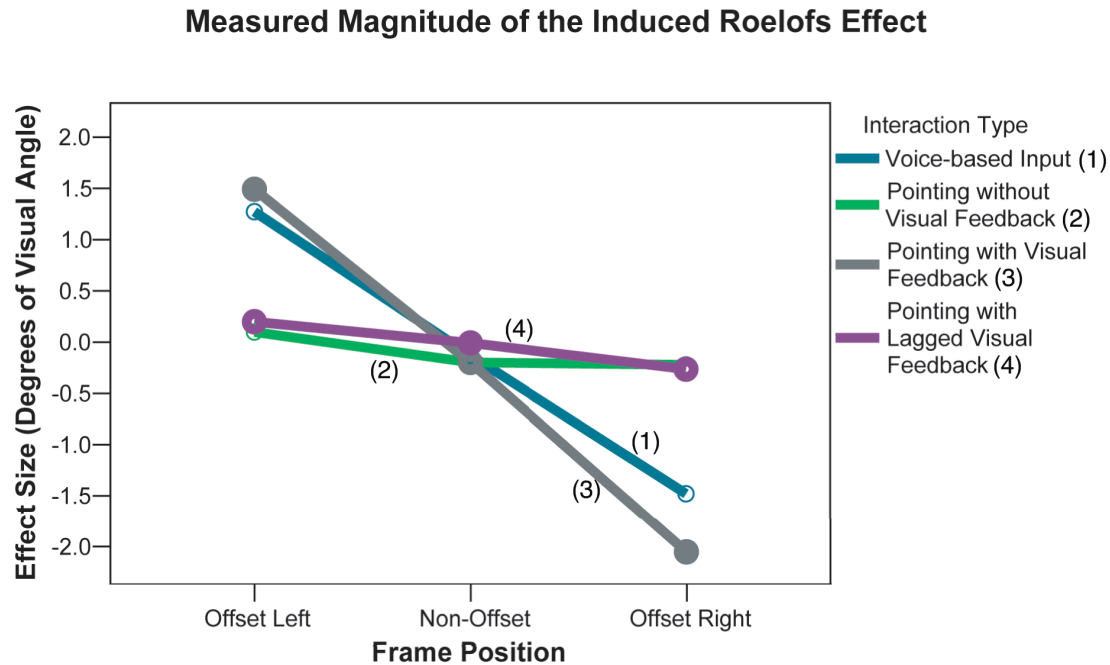


Figure 7.5: A marginal means plot measuring the induced Roelofs Effect across the different interaction types and varying frame positions in this experiment. Effect size indicates the degree to which participant responses deviated from actual target positions. “Negative” effect sizes indicate deviations to the left while “positive” effect sizes indicate deviations to the right, measured in degrees of visual angle. The steep slopes associated with (1) voice-based input and (3) pointing with visual feedback indicate these interaction types were highly susceptible to the induced Roelofs Effect. The corresponding horizontal slopes associated with (2) pointing without visual feedback and (4) pointing with lagged visual feedback demonstrate that these interaction types were substantially less affected by the visual illusion.

they exhibited evidence of an induced Roelofs Effect). The figure presents a marginal means plot of aggregate participant performance as assessed by the one-way ANOVA. The steeper slopes associated with the voice-based input and pointing with visual feedback conditions characterize a systematic bias in participant response associated with frame position, as would be expected in the presence of an induced Roelofs Effect.

#### 7.4.5 Performance with Voice Input

Consistent with the experimental hypotheses, voice-based input was found to have the highest degree of overall susceptibility to the induced Roelofs Effect. Based on the aggregate analysis, a statistically significant main effect of frame position was found overall for voice-

---

based responses [ $F(2, 766) = 252.85, p < .001$ ]. This concurred with the individual analyses, which found that sixteen of the twenty-four participants had significant main effects of frame position [ $F(2, 30) > 3.35, p < .049$ ]. The size of the bias is consistent with previous experiments of the two-visual systems hypothesis, which found that the response bias was approximately 1.5 degrees of visual angle.

#### 7.4.6 Pointing with Visual Feedback

Pointing with visual feedback was found to be the input method with the second-highest degree of overall susceptibility to the induced Roelofs Effect. The aggregate one-way ANOVA found a significant main effect of frame position [ $F(2, 766) = 27.91, p < .01$ ]. The individual analyses found that fourteen of the twenty-four participants had significant main effects of frame position [ $F(2, 30) > 3.80, p < .034$ ]. The size of the response bias was found to be somewhat greater when the frame was offset to the right (2.0 degrees to the right) than when the frame was offset to the left (1.5 degrees to the left), although the overall response bias remains consistent with the induced Roelofs Effect.

#### 7.4.7 Pointing with Lagged Visual Feedback

Pointing with lagged visual feedback appeared to have substantially less susceptibility to the induced Roelofs Effect. Consistent with the experimental hypotheses, no evidence for the presence of the visual illusion was found. The aggregate analysis found no overall statistically significant main effect of frame position for pointing with a lagged cursor [ $F(2, 766) = 2.26, p = .105$ ]. This was also consistent with the individual analyses, where none of the participants were observed to have significant main effects of frame position individually [ $F(2, 30) < 1.82, p > .180$ ]. Overall response bias was measured to be less than 0.25 degrees of visual angle, which could be conceivably attributed to the errors implicit in making pointing movements.

#### 7.4.8 Pointing without Visual Feedback

Pointing without visual feedback demonstrated results almost identical to those found in pointing with lagged visual feedback. No evidence for the presence of the induced Roelofs Effect was found. The aggregate one-way ANOVA found no overall significant main effect of frame position when pointing without any visible cursor [ $F(2, 766) = 1.29, p = .277$ ].

Furthermore, none of the participants were observed to have individually significant main effects of frame position [ $F(2, 30) < 0.88, p > .425$ ]. The size of the response bias was less than 0.25 degrees of visual angle, as was also observed when pointing with lagged visual feedback.

## 7.5 Discussion of Experiment 3

The results derived from this experiment provide evidence in support of the outlined experimental hypotheses. Voice-based input and pointing with visual feedback appear to exhibit performance characteristics consistent with responses that draw upon a ventral (perceptual) representation of visual space. Pointing without visual feedback and pointing with lagged visual feedback appear to have been relatively immune to the effects of the induced Roelofs Effect, suggesting that these two methods of interaction may be more likely to draw upon a dorsal (motor) representation of visual space. These results collectively suggest that visual illusions like the induced Roelofs Effect could affect user performance in more practical situations, especially where large screens and multimodal input play a role in user interaction.

Compared to other experiments involving the two-visual systems hypothesis, the results remain fairly consistent. In other experiments, such as those conducted by Bridgeman, Peery, and Anand (1997), it was found that some participants did not appear to be affected by the perceptual biases of visual illusions regardless of response method and the same result was found in this experiment (only sixteen of the twenty-four participants exhibited an induced Roelofs Effect with voice-based input). A possible explanation for this, as in these earlier studies, is that some individuals are more inclined to draw from a dorsal (motor) representation of space, even for certain kinds of perceptual identification tasks. This result could therefore be the result of characteristic individual differences between the participants who volunteered in this experiment.

This experiment has demonstrated that there are many assumptions usually made about the kinds of elements that should be present in user interaction, and that some of these assumptions might not always hold true. In particular, the presence of systematic response errors when pointing with visual feedback versus pointing without a visible cursor suggest that it may be possible to devise methods of input robust against perceptual ambiguities like

---

visual illusions in large screen display settings. Although techniques like pointing without visual feedback may currently suffer from problems such as increased variation in user performance, it is conceivable that it may one day be possible to characterize such variation as a function of the individual as a means of compensating for pointing error. The ability to reliably point without visual feedback would be valuable in many circumstances where visual cursors are deemed “necessary,” including synchronous, collaborative applications where the presence of multiple cursor representations might be distracting or might depend on increased effort on the part of users.

The question of what visual elements flag the mental representation of a visual task as either perceptual or motor is partially addressed in the current experiment, with its emphasis on the presence or absence of visual feedback for pointing. Other experiments in the future could look at various other visual elements on a graphical display, including the perceptual biases that might be associated with other kinds of framing elements like lines, scales, and shadows (in 3D environments). Although these might not seem particularly important in desktop environments, the relative lack of physical or graphical context cues with large screen configurations could mean that these kinds of graphical elements might take on stronger, and perhaps unanticipated, roles in these settings.

### 7.5.1 Implications for User Interface Design

Although the experimental work presented here may appear to be inherently theoretical, the results suggest that the mediation of visual information as described by the two-visual systems hypothesis has value in the study of HCI. The following are some of the more practical lessons to be learned from this investigation:

- *The relationship between visual perception and motor action could be important to the study of HCI.* There is increased interest in large screen display environments and the emergence of multimodal interaction techniques. Unlike desktop settings, where there are numerous kinds of perceptual framing cues like the physical edges of a monitor, large screen displays are more immersive in the sense that the only kinds of contextual cues presented to users may be those provided by the graphical interface. This could lead to an increased chance that perceptually ambiguous effects like visual illusions might have an impact on user performance. This further suggests that the

way in which visual elements are deployed for smaller screens may not necessarily scale upward to larger screens with the same kind of perceptual tolerances.

- *Perceptual judgments are not necessarily the same as motor judgments.* The ability to judge object sizes and spatial locations is one that has been important to the evolutionary development of human beings as a species, and it remains one that is important for a wide variety of computer-related tasks such as computer-aided design (CAD) and design reviews in engineering. For systems that contain time- or safety-critical elements, it is possible to guard against perceptual biases by using motor interactions and visual characteristics that are more likely to draw upon the dorsal (motor) representation of visual space over the ventral (perceptual) representation of visual space.
- *Voice interaction is more reliant on perception than is gestural interaction.* This point follows as a consequence of the previous point. The predictions arising from the two-visual systems hypothesis suggest that effective voice input will require designers to be more careful about avoiding perceptual ambiguities in the visual structure of display information. These might range from basic visual characteristics such as colour or texture, but might also include elements such as spatial relationships between visual objects. Including visual elements that encourage motor behaviour might be valuable where motor behaviour is likely to reduce effort on the part of users or might help users avoid predicted, unintentional errors of execution.
- *Even basic graphical elements can have an impact on visually-guided interactions.* This experiment has demonstrated how even simple graphical elements like rendered frames and visual cursors can bias user performance. Thus, seemingly “obvious” design choices like the inclusion of a tracking cursor or the presence of contextual asymmetry should be carefully assessed in certain scenarios. The two-visual systems hypothesis indicates that the mechanisms enabling visually-guided interactions are not especially intuitive because visual processing occurs at an unconscious level. This could be especially important when graphical elements are placed in the context of a much more complex display and cognitive resources are particularly limited. The results of this experiment suggest that the minimization of visual information could be used to learn how to make interaction less demanding for users.

## Chapter 8

# Discussion and Applications

The previous three chapters have presented three different experiments motivated by the integration (S-R compatibility), separation (the functional specialization of the UVF and LVF), and mediation (the two-visual systems hypothesis) of visual information. These experiments have demonstrated how a representational approach to studying HCI could be valuable in identifying systematic variations in user performance that are not explained by existing theoretical frameworks. The implications for user interface design derived from the results of these three experiments in the previous chapters demonstrate how a representational approach to HCI might translate into valuable lessons for practitioners who may be less concerned with their theoretical implications, but may be more interested in the practical impact. This chapter expands upon the implications discussed earlier by further arguing how and why a representational approach to HCI can be valuable for the design and evaluation of interactive systems.

### 8.1 Theoretical Frameworks in HCI

Perhaps the most valuable way in which a representational approach can be beneficial to the study of HCI lies in the way it affects the role of psychology in the discipline. Historically, HCI has been an area of study that is motivated by the techniques, tools, and knowledge available in the psychological disciplines. Psychology and related fields, such as kinesiology, have provided a number of theoretical frameworks such as Fitts's Law and GOMS, which have proven useful in the design and evaluation of interactive systems. The relationship between psychology and HCI has become somewhat tenuous in more recent years, as newer technology that does not have counterparts in traditional work or play settings has become available. There is a great demand for new applications that make use of these new kinds of interactive tools, but seemingly less knowledge to draw upon in the literature. This is not to suggest that psychology no longer has a role in HCI, but perhaps that the relationship

between the study of human behaviour and HCI has become less evident in recent years.

With the volume of literature related to controlled, experimental evaluations of user performance available jointly in the areas of HCI, ergonomics, and experimental psychology, it could be argued that HCI may benefit more from an emphasis on the development of practical designs and methodologies rather than a continued focus on theoretical results. The results derived from the representational themes examined here should nevertheless demonstrate that there continue to be very good reasons to continue these kinds of more basic, theoretical work in the study of HCI. Because existing approaches to evaluation in HCI would not have directly predicted the user performance variations already well-known in cognitive psychology, there are likely many other as-yet unexplored instances where user performance could be systematically manipulated by altering one, or a few, basic characteristics of an interface as a consequence of the way that users process their interactions with a computing system. Identifying these characteristics and understanding the ways in which they can be manipulated to achieve intended user performance is becoming more valuable as user interfaces continue to grow increasingly complex. Such an understanding will only become more valuable in the future because increased complexity in the interface leads to increased demand for attentional resources, indicating that there is a need to develop alternative ways in which users can be informed about the state of a system without incurring additional cognitive load.

Thus, a representational approach to HCI could serve as the foundation for a new thread of evaluative research, tightly coupled to the psychological basis from which HCI initially arose, but with a “modern” emphasis on currently relevant tools, techniques, and applications. The emphasis on mental representations here is also congruent with the advances in cognitive psychology since the early Model Human Processor work of Card, Moran, and Newell (1983). In their experimental work, the importance of mental representations had yet to achieve the impact that it currently has on the study of human behaviour. The emphasis on mental representations of perception and action discussed in this dissertation is largely motivated by interpretations of psychological evidence that have only arisen in recent years. Since mental representations have been, and will continue to be, fruitful ground for psychological research, it seems only fitting that HCI might do well to start exploring a shift from a purely descriptive approach to understanding user performance to one that also *explains* the factors that drive user performance. At the very least, a theoretical framework



---

based on mental representations can predict the conditions under which certain behavioural phenomena may or may not affect user performance, and therefore when they should be a concern for interface designers.

The results derived from the three experiments presented here may also serve as evidence that there are many concrete application areas for representational theories from cognitive psychology. Even theories that have an established history in HCI, such as S-R compatibility, could continue to provide “new” insight into problems arising from existing and future kinds of interactive techniques and technology. The experiment comparing cursor orientations (Chapter 5) presented here is but one example of an issue that has long gone unstudied in HCI, and yet reveals results that are “surprising” in the absence of psychological theory. The UVF and LVF (Chapter 6) and the two-visual systems hypothesis (Chapter 7) have looked at basic issues in user interface design, including spatial organization and the presence of visual cursors respectively. These two particular threads of representational study could be valuable in the context of ubiquitous computing and multimodal interaction, where new technology is being deployed at a rapid pace, but there remain many open questions about how useful, seamless interaction can actually be achieved in practice.

## 8.2 Developing New Interaction Techniques

A representational approach to HCI need not only be important from the perspective of experimental evaluation. It is conceivable that the representational theories presented here, and the many aspects of mental representations that have not been explored in this dissertation, could be helpful in inspiring new kinds of interaction techniques in a wide variety of settings. For example, Rensink (2002b) has suggested that new techniques based on the representational aspects of attentional coercion and non-attentional pickup could be used to control the behavioural systems that allocate visual attention, preventing costly mistakes from being made and creating an interactive work environment that minimizes disruption in the workflow of users.

Each of the three representational experiments leave open questions about how the derived experimental results might be exploited as new methods for user interaction. In the study of cursor orientations and S-R compatibility, it was suggested that a real-time, dynamic pointing cursor could theoretically improve pointing performance by maintaining

S-R compatibility, regardless of movement direction. In the study of mouse and touchscreen pointing in the UVF and LVF, it was suggested that dynamically changing the spatial layout of an interface to continually keep the most relevant interactive elements in the LVF might theoretically improve user motor performance. In the study of pointing, visual feedback, and the two-visual systems, it was suggested that the presence or absence of visual cursors could be exploited to make interaction in busy, collaborative computing applications less demanding on the part of users.

Designer experience would probably suggest that the purely theoretical performance gains suggested by the results of the three experiments presented here are impossible to attain by simply implementing these suggested schemes. This is not because the experimental inferences drawn are incorrect, but simply because the effective performance gains achieved by putting these design suggestions into practice would undoubtedly be mediated by factors untested in the current experiments, including the visual context of the user interface, other perceptual considerations like attentional focus, and the end goal of the user interaction. Nevertheless, these design ideas could form the conceptual basis for arguably more “practical” interaction solutions based on the specific application issue being addressed.

One particularly interesting avenue of future design work related to the representational theories presented here involves the concept of individual variability, which has been touched upon in the data analysis of the three presented experiments. In HCI, there remains a tendency to develop interfaces for aggregate performance while ignoring the fact that no single user will ever precisely conform to an aggregate model. This is in contrast to much of the research that is done in cognitive psychology, and increasingly in HCI-related fields such as ergonomics, where individual differences are rapidly becoming a topic of substantial interest. Representational theories may be of particular interest to designers of adaptable or adaptive user interfaces because these theories are now starting to incorporate explanations for why these kinds of differences occur and how these differences manifest themselves in the “outside” world. The individual variability seen in experiments such as those by Bridgeman, Peery, and Anand (1997) in the context of the two-visual systems hypothesis provide a good example of this trend (see Section 7.5 for more discussion). Thus, an investigation of representational theories for individual variability among users might be helpful in coming up with concrete ideas about the ways in which an interface could be customized to support individual users.

---

At the greatest extreme, the psychophysical techniques employed to study the representational theories may eventually give way to a “personal equation” of users akin to the concept of adjusting for individual variability to account for systematic errors of observation in astronomy. The term “personal equation” can be attributed to Bessel in 1822, with subsequent work by Wolf (1865). They found that different astronomers made some kinds of observational measurements consistently differently. Wolf showed that the characteristic variance between measurements could be minimized by generating a unique equation for each individual, thereby improving the consistency, precision, and overall accuracy of the measurements. In HCI, a personal equation of interaction with very similar benefits may be possible in the future. The goals for research into such an equation would include the identification of major parameters involved in user interaction and what the appropriate measurements would be for particular individuals.

With the appropriate technology, such a personal equation might be able to address “larger” individual differences when working with a visually-complex user interface, such as colour blindness or binocular (depth) acuity, and less evident issues involving individual variability, such as pointing accuracy. These are aspects of interface design that are becoming increasingly important as the development of “natural” or “seamless” interaction becomes a focal point for HCI research. This may also be of interest in collaborative situations where it is important for everyone to “see” the same thing, although different participants may have different system configurations, as is the case when some users may be working in a large-screen environment, and others may be remotely working from a laptop or smaller, portable displays.

### 8.3 Ubiquity, Immersion, and Presence

A problem of some widespread interest in HCI is understanding the value of large screen displays and devising objective measures of “immersion” or presence. Czerwinski et al. (2003) point out that despite the numerous qualitative claims about the benefits of immersive display technology, few empirical investigations in the literature demonstrate real or perceived productivity benefits from using multiple displays or large displays. Although there exist standardized tests to quantify certain aspects of immersion, these generally rely on the subjective impressions of those taking these tests and they fail to address the role of

non-conscious processing in determining user behaviour in immersive settings, even though such processing may be very important. Because large screen and immersive environments will continue to be critical in many kinds of interactive scenarios, it has become important to objectively characterize the situations in which immersion or presence might be valuable, and to identify the visual characteristics of a display setting that “increase” a user’s sense of immersion or presence.

These are also issues that are important in the realm of ubiquitous computing, as the distinction between physical and virtual (i.e. computer-generated) objects begins to merge. Multimodal interaction and the interest in tangible user interfaces has led to the rapid development of new technology, software toolkits, and many promising ideas about the way that users might interact with computers in the future. However, there remains very few reliable methods for evaluating the impact or performance effects of such technology. Many evaluations reported in the HCI literature revolve around the same kind of dependent measures used in the three experiments of this dissertation. However, many lack a substantial theoretical grounding from previous psychological work, and rely on “designer intuition” to make the argument that certain ideas and implementations have merit.

Theories of mental representation, including the ones studied here, may be especially valuable in addressing these particular problems. The wide breadth of knowledge about visual perception and the processing of visual information as studied in vision science and cognitive psychology could hold valuable clues as to when large screen displays are beneficial and when they may be harmful. For example, in the previous discussion of the two-visual systems hypothesis, it was demonstrated that certain kinds of visual illusions such as the induced Roelofs Effect, could appear in these kinds of display environments and that avoidance of such perceptual biases should be avoided when minimization of user error is a crucial design goal. Other psychological research surrounding the topic of visual cognition, such as that of the separation between peripersonal and extrapersonal space, could prove invaluable in building a basic understanding of how people perceive the visual space of a large screen or immersive environment and how these kinds of particular perceptual differences might be exploited for the benefit of the user.

The evidence from the other two investigated representational themes (S-R compatibility and the distinction between the UVF and LVF) provide concrete examples of how such research might be achieved. With respect to S-R compatibility, the literature suggested that

---

different hand postures could influence the compatibility of user input to different degrees and this was borne out in the collected experimental data. A representational theory such as S-R compatibility might be particularly important when designing interactive applications that must span small, medium, and large displays because it provides a uniform theoretical basis for making predictions about user performance across all different display sizes. With respect to the UVF and LVF, the literature suggested that the UVF and LVF might be functionally specialized for processing information about extrapersonal and peripersonal space respectively and this was also seen in the experimental data, where a preference for motor interaction was found in the LVF. A representational theory such as that encompassed by the UVF and LVF might be important to understanding interaction in ubiquitous settings because the theory places an emphasis on how users' perception of the surrounding world is changed when interacting "up-close" versus interacting "at a distance."

## 8.4 Application Domains

Based on the experimental evidence derived from the three representational themes explored here, there are at least four application areas of interactive system design that might specifically benefit from a representational approach to HCI. These areas are broadly defined as (1) time- and safety-critical systems, (2) computer graphics, (3) information visualization, and (4) computer-supported cooperative work. In each of these application domains, the themes of S-R compatibility, the division of the UVF and LVF, and the two-visual systems hypothesis relate knowledge that could be particularly valuable to the design and evaluation of systems involving these applications. Where applicable, the importance of other areas related to the study of mental representations is also discussed.

As it currently stands, the design of practical interactive systems generally assumes that the visual world of users can be described as a unified percept. The experimental evidence from representational theories of visual space disagree, and the three experiments presented here provide additional evidence that it remains important to challenge the notion that the visual world is intuitively "what we see" and no more than that. Although it is clear that more research needs to be done to fully understand the potential benefits of applying these representational theories of visual processing to user interface design, the basic experiments presented in this dissertation suggest that enhancing user performance

does not necessarily mean that designers must completely change the way in which users interact with a computer. Rather, measurable performance gains can be made by simply paying more attention to some of the more basic visual elements of a user interface, such as the directional cues of a visual cursor, the spatial layout and organization of an interface, and the presentation of appropriate visual feedback.

### 8.4.1 Time- and Safety-Critical Systems

Time- and safety-critical systems refer to those interactive systems where user response time, or the safety of someone involved with an interactive system are crucial design considerations. These range from the design of vehicles, such as cars or aircraft, as well as medical care systems, such as computer-assisted surgery systems. Many of the primary results from the three presented experiments are directly applicable to the development of these systems. The systematic variations in user performance predicted by the three representational theme manifest themselves most measurably as differences in response time and participant accuracy. Where time is a critical factor, every millisecond that can be saved is important. If we use driving as an example, a vehicle traveling at 50 kilometers per hour moves approximately 1.4 meters every 100 milliseconds: a substantial distance when we consider the kinds of interactive tasks that might need to be accomplished while driving.

The three representational themes demonstrate that there are some basic aspects of the visual interface that can be improved upon by simply making an effort to understand the different ways in which users integrate, separate, and mediate visual information via multiple representations of visual space. A similar point can be made where safety is a critical factor. In safety-critical systems, minimization of user errors are important not only because errors cost time, but also because errors are can be irreversible. With respect to pointing, the three experiments presented here demonstrate that improvements in response time are generally accompanied by improvements in user precision.

S-R compatibility suggests that the presence of directional cues of any kind can have an implicit impact on user performance. The measurable effects of co-varying arrowhead cursors and movement directions in the current experiment suggests that directional cues must be employed strategically and should not be used where they are not especially necessary. To facilitate user performance in a system based on pointing interactions where no movement cursor is preferred, an orientation-neutral cursor might be best. These sugges-

---

tions may extend to systems where many different kinds of directional movements must be made concurrently and rapidly, and not all of these can be strictly characterized as pointing. Driving a vehicle is again one of the better examples: use of a steering wheel implies that certain directional judgments must be made, and often times while driving, movements of the steering wheel are interspersed with pointing movements, such as turning a car radio on or off by pressing a button. Poorly placed directional cues could be confusing or dangerous to the driver, and S-R compatibility suggests that the presence of multiple, conflicting directional cues could severely impede movement performance.

The presence of a functional division between the UVF and LVF suggest that motor performance is superior when items are initially perceived in the LVF. For systems where user response time is especially critical, this suggests that the most important interactive elements should be placed where they are most likely to be initially seen in the lower half of the visual world. This might further suggest placing such items in the lower half of a display or interface. Because the results suggested by the UVF/LVF experiment presented here are consistent with other experimental evidence related to the visual fields, it is reasonable to expect that similar performance benefits could arise by employing the functional specialization of the UVF appropriately as well. The literature suggests that the UVF is specialized for perceptual activities, and as such, interfaces that may include rapid changes to the visual display could do well to consider placing the most important perceptual elements such that they are most likely to be perceived in the UVF.

The two-visual systems hypothesis indicates that certain representations of visual space can be biased by the presence of visual ambiguities, like visual illusions. For time- and safety-critical systems, the hypothesis suggests that these illusions can implicitly impact performance under certain conditions. Perhaps most important for these systems is that these perceptual biases can go unnoticed by the user, and the user may believe that he or she has not committed an error. The experimental results indicate that the perceptual biases induced by visual illusions can be noticed in a large screen display setting and that robustness against these kinds of biases are the result of input method and the presence or absence of visual feedback. Pointing without visual feedback draws from the dorsal (or unaffected) representation of visual space while more intuitively reliable techniques, such as pointing with a visible cursor or voice-based input draw from the ventral (or affected) representation of visual space. When designing a system that seeks to minimize user er-

ror, the two-visual systems hypothesis generally suggests that it is important to identify those visual structures that might be perceived as visual illusions, as well as those visual characteristics that might bias users to process visual information in unintended ways.

### 8.4.2 Interaction with Computer Graphics

The development of new interactive techniques has historically been an important part of computer graphics research. As HCI has grown beyond evaluation to encompass design over the years, the emphasis in computer graphics has shifted away from developing new techniques involving a relationship between vision and movement and more toward the development of techniques that are purely perceptual in nature. Nevertheless, a representational approach to HCI may be beneficial to understanding the human perceptual issues surrounding the study of computer graphics. Insofar as human judgments are necessary to make these graphics useful and because of the insight that mental representations have helped to generate in cognitive psychology and vision science, there may be many ways in which the knowledge from a representational approach to HCI could be valuable.

There is a growing interest in computer graphics toward perceptually-based rendering techniques, or methods for generating computer graphics that are based on the knowledge of how users process visual information. Instead of the more traditional brute-force approach to rendering, involving massive amounts of computation that still cannot be achieved in real-time, perceptually-based rendering often takes a more elegant approach of only rendering what is “important” to the user. In this sense, “important” could mean where users have directed their gaze, or perhaps what the most salient aspects of the visual display are likely to be. Mental representations related to visual processing could be especially valuable here, as they continue to characterize the phenomenon of visual attention and the mechanisms that allow human beings to make reasonable judgments in what might otherwise be a perceptually overwhelming world. A study by Harrison, Rensink, and van de Panne (2004) provides an example of where the implications of such a model of visual processing have proven to be useful. They found that length changes of over twenty percent could go unnoticed in the manipulation of an animated, articulated figure when the figure was not given full attention, yielding guidelines for optimizing attention-based algorithms that produce or process character motion data.

The functional distinction that arises from the UVF and LVF serves as a good example



---

of a representational theme that could be valuable in this context. The UVF is largely specialized for perceptual activities, suggesting that perhaps with additional research, new adaptive rendering techniques could be devised that exploit this particular functional advantage. Such an algorithm may choose to place priority on rendering the upper half of a visual scene before rendering the lower half. Experimental evidence with respect to the visual fields also suggests that there is greater attentional resolution in the LVF, suggesting that other algorithms might explore the employment of rendering strategies that involve generally coarse rendering of a visual scene augmented by finer details that are most likely to be perceived in the lower half of the visual world.

At least partially related to the study of computer graphics is the design and development of new techniques in support of computer gaming and entertainment. While gaming is not generally the focus of pure research, it is often cited as a potential application area for interesting results (O'Sullivan & Dingliana, 2001; O'Sullivan, Dingliana, Giang, & Kaiser, 2003). A particularly interesting application for mental representations might involve the characterization of visual gaming stimuli and the kinds of emotional responses they evoke. Such work is already underway in certain parts of experimental psychology, and the results of such research could be used to understand the elements of particular gaming genres that make them appealing to particular audiences. Such research could also be valuable in streamlining the user interface designs that accompany computer games by helping designers understand what basic visual aspects of a gaming interface are most likely to make users confused or frustrated.

### 8.4.3 Information Visualization

The application area of information visualization is unique because it lies at the crossroads between computer graphics and HCI. Thus, there are many perceptual aspects of real-time rendering that are important to information visualization applications, but the element of interactivity remains crucial as well. The goal of many information visualization applications revolves around the visual management of very large and potentially very complex sets of data. Among other applications, such data may include the presentation of information for the analysis of DNA and genetic sequences, or it may involve the visualization of higher-dimensional data for scientific analysis. Where the design of tools that support the visual management of large amounts of data is concerned, the study of mental repre-

sentations in HCI could yield valuable clues about how such tools should be designed, as is described by Ware (2004) in his overview of the basic characteristics of visual processing and its relationship to information visualization.

Arguably, one of the major reasons that so much of the human brain is devoted to the task of visual processing is because the visual world is extraordinarily complex and using a brute-force approach to comprehending all of this information would be largely infeasible. Thus, very complex mechanisms (such as the division of the vertical visual fields and the ventral and dorsal streams of visual processing identified in the two-visual systems hypothesis) are necessary to determine what aspects of the visual world are most relevant at a given moment in time. This is consistent with the goals of information visualization tools, which must find ways to cope with information that might not otherwise be easily understandable. Although the mechanisms of human vision have not specifically evolved to deal with processing the kinds of visual data seen in these applications, an understanding of mental representations and their implications for visual processing could be helpful.

Both S-R compatibility and the UVF/LVF provide insight into how visual information can be managed and how interaction can be used to support the management process. Since many information visualization applications involve the directional scaling and rotations of spatial data (i.e. as might be found in the analytical maps of geographic information systems), S-R compatibility could help designers learn about what visual cues and interactive movements are appropriate to efficiently navigate through large data sets. They may also be valuable in helping designers determine what kind of movements and associated responses could be confusing to users ahead of time (i.e. deciding which directional movement of an input device might best be used to rotate a complex 3D visual structure where several possibilities are open to designers). The knowledge gained from the UVF/LVF follows from the previous discussion in computer graphics. The functional specialization inherent to perceptual tasks in the UVF could be used to help decide which aspects of an information visualization should be rendered first. The UVF/LVF distinction could also be used as the basis for new focus+context visualization methods, perhaps where the UVF is used to provide context, but the LVF is used for focus and finer interactive manipulation.

A particular challenge to research in information visualization also revolves around the lack of evaluation in the area. Much research involves the development of new systems or new visualization techniques without a substantial component of accompanying evaluation.

---

A representational approach to HCI could be quite valuable in addressing this problem by offering a novel theoretical grounding for understanding why and how certain systems are effective while others are not. The emphasis on basic visual characteristics presented here, as well as the study of visual cognition at large in other related research in cognitive psychology could be used to study the visual elements common to information visualization systems and how these might be used to help develop scientific theories specific to the application domain. In particular, the focus on mental representations could be helpful in getting the designers of such systems away from simply thinking about the technology supporting the visualization and toward thinking about the users who must make use of the end visualization.

#### 8.4.4 Computer-Supported Cooperative Work

With the rapid deployment of large screen displays and the emergence of ubiquitous computing, the idea that computing environments can be useful to support the concurrent work activities of multiple users has gained momentum in recent years. The area of computer-supported cooperative work (CSCW) looks at the ways in which computers can be used to help people work together, regardless of whether they are all in the same room or are separated geographically but connected via a communications network. Though the experimental work presented here has consisted solely of studies in the context of single users, a representational approach to HCI may also be useful in CSCW.

From an organizational perspective, the basic lessons learned from the representational themes explored in this dissertation continue to apply. S-R compatibility suggests that directional compatibility is not simply the result of what is shown on the display and how it relates to the use of a particular input device. It is also the result of other factors, including user orientation and the spatial layout of the environment. Knowledge of S-R compatibility may be especially important in situations where screens and users are distributed across an entire room, or perhaps when people are working around horizontally-oriented configurations like tabletop displays. This could suggest the implementation of a mechanism for adaptively altering input compatibility based on a user's position in a collaborative environment, since a visual item that may be on one user's left may be on another user's right and the compatibility of user input when interacting with this item could potentially vary.

The experimental work involving the two-visual systems hypothesis could also be of some interest here. The hypothesis suggests that visual feedback is not always necessary, and that there are in fact situations where it is detrimental to user performance. There currently exist software toolkits and frameworks for enabling collaborative applications with multiple cursor representations and synchronous collaboration with input devices like multiple mice or laser pointers. The full impact of these “new” methods of interaction has yet to be assessed, as relatively little research has been done to explore how such collaborative interaction could be best employed. One particular challenge involving multiple cursors involves assigning unique identifiers that can be perceptually identified with specific users. In environments with many potential collaborators (i.e. an interactive classroom), it may be infeasible to assign a unique cursor representation to each user, or it may simply make the collaborative workspace unusable as it becomes cluttered with cursors that obscure the interface or the information that might be relevant for the end task. Knowing that visual feedback as suggested by the two-visual systems is not always necessary may be an important clue in addressing an issue such as this. In future systems, it might be possible to eliminate visual feedback for pointing completely or it may be possible to limit visual feedback to those users who need it for very specific kinds of system interaction.

The study of mental representations for encapsulating group knowledge is also central to other aspects of cognitive psychology, including the study of distributed cognition, which proposes that cognition and representations of knowledge are not bounded to a single user, but are rather distributed across individuals and tools in the surrounding environment. Thus, the judgments made by a group of users working together are more than simply summing together the outputs of individual users. Though this particular aspect of knowledge representation was not explored here, the representational approach to HCI that has been presented may yield insight into how studying these theories of group representation might be possible in a systematic fashion. It may also be important in devising new methods for understanding user interaction in groups or the creation of effective design methodologies for collaborative systems in the future.

---

## 8.5 Other Issues and Limitations

Having presented the three representational themes that form the bulk of the evidence supporting the central argument of this dissertation, a few words should be said to compliment the comments made in the Introduction (Chapter 1) with respect to the scope and objectives of this thesis. The representational theories presented here are not meant to exclude the implications of other bodies of theory in HCI, many of which may not be exclusive to the study of cognitive psychology. One of the goals of applying representational theories to HCI involves complimenting existing frameworks for evaluating interactive tools and techniques rather than seeking to replace them entirely. Although the central thesis makes the claim that particular aspects of user performance have previously gone unnoticed, other evaluation methods have been shown to be equally useful in providing insight that cannot be found through an application of representational theories alone. Thus, the representational approach to HCI as presented here, should be interpreted as an additional tool now available to researchers and practitioners to better understand users rather than a tool whose goal is to deny the value of existing work.

The psychophysical methods employed in this dissertation have been central to demonstrating that systematic variations in user performance exist in a manner consistent with the predictions of representational theories. By choosing to use a more controlled, experimental approach to evaluation, it can be rather difficult to understand how these results might be put into practice. This could be cited as the primary limitation of the representational approach as presented here: the performance variations uncovered by experimental work are quite visible in a laboratory setting, but there are few direct examples of how such theories have influenced the design or evaluation of actual technology in practical use. Nevertheless, the particular results derived from the three investigated representational themes are amenable to immediate practical application. The experiments have emphasized the importance of basic visual and input characteristics as opposed to the limitations of any particular technology. Furthermore, the implications for interface design derived from these experiments were presented in a generalized manner so that their importance would not become tied down to only one or a few different application areas. When presented in this way, the results of the experimental work presented here become insight helpful to HCI generally rather than becoming very specific prescriptions for the way in which user interfaces

must be designed.

Another major issue related to the practical application of the results presented here revolves around the issue of small effect sizes. In the experimental work presented here, performance differences were measured in relatively small units, such as milliseconds or degrees of visual angle. It could be argued that these small effect sizes alone make the work impractical, since the particular performance savings are negligible for any particular interaction. There are at least two reasons why such performance savings are valuable, regardless of how small they might seem. First, when one considers how important pointing is as a means for interaction, the cumulative benefits of saving several milliseconds with each pointing movement becomes substantial over time. Pointing will likely continue to be an important aspect of user interaction in the foreseeable future, and with the increasing use of less steady input devices like wands and pointers in a variety of situations, identifying methods for improving the efficiency of such input devices becomes critical in minimizing user confusion and frustration. Second, there is a growing body of work in cognitive psychology related to understanding the “flow” of human behaviour. With a state of flow comes enhanced focus and overall productivity in a given task, which suggests that understanding flow might be an important research topic in HCI. Where small performance benefits are concerned, it is plausible that destroying a user’s sense of flow could happen on the order of a few milliseconds, whether it be because of an unexpected visual event (i.e. a sudden change in the graphical display), or because a pointing interaction required additional cognitive effort on the part of a user (i.e. more time was needed to configure the appropriate pointing movement). Thus, optimizing performance by saving users a few milliseconds could be crucial by indirectly preserving a user’s state of flow during interaction.

Several major assumptions have been made with respect to the application of representational theories here. The largest of these was outlined earlier, with respect to computational perspectives on the nature of the human mind, and its relationship to the human brain. This thesis generally assumes that the human mind can be described as a computational entity and that the plausibility of this assumption is derived from previous evidence in cognitive psychology, the impact that applying mental representations has had on understanding human behaviour, and the strong historical influence that information processing frameworks have had on the study of HCI. It is useful to reiterate that this assumption does not exclusively deny the existence of other bodies of psychological theory, nor is it the

---

intention of this thesis to do so. Other approaches to understanding human behaviour may prove valuable to HCI and this may be demonstrated by others in the future. However, the major goal of this thesis has not been to argue whether the mind can be described in any particular way, but rather to demonstrate how one particular description can yield new insight to understanding user performance.

Another assumption related to the investigated representational themes is the presence of multiple, concurrent representations of visual space. Early work into computational theories almost exclusively assumed that a given mental representation had a single symbolic output. From a philosophical standpoint, it may be difficult to determine whether the resulting differential outputs of a given visual input (i.e. the separable mental representations for perception and action) are the result of two separate mental representations, or that of a single representation with two possible outputs. This particular assumption makes it especially evident that it is important to keep up with current trends in cognitive psychology, as new theories are constantly presented that supercede previous beliefs. With respect to this particular issue, there is substantial evidence that there are multiple, separate representations as opposed to a single, complex representation that mediates visual behaviour. The neuroanatomical descriptions that underlie the functional specialization of the UVF and LVF, as well as the presence of ventral and dorsal streams of visual processing as described by the two-visual systems hypothesis serve as physical evidence that separate representations for perception and action exist. The “double dissociation” paradigm used in cognitive psychology to experimentally demonstrate systematic differences in visual behaviour (i.e. as used to study the two-visual systems hypothesis) also serves as further evidence that the observed behavioural outputs are not simply the result of a single mental representation.

## 8.6 Future Work

The representational approach to HCI and the three themes presented in this dissertation open up the possibility for several promising directions of future research. Perhaps most obvious is the actual application of the results presented here, and comparing the “real-world” impact on user performance relative to the theoretical data that have been collected. Closely related is the design of new interaction techniques that augment existing ones, motivated by the lessons learned from fostering a representational understanding of user behaviour.

In addition to studying specific cases where the lessons learned from the representational themes are applied to design and evaluation of real systems, there are several directions for further evaluative research within each of the three themes.

The Simon Effect is a phenomenon closely related to the study of S-R compatibility, and is already a thread of valuable research in cognitive psychology. Since the effect suggests that compatibility effects can occur even when they are not relevant to the task, this suggests that there may be instances where S-R compatibility influences user performance even though it may not be intuitively apparent that S-R compatibility is important. Synchronous display onset asynchrony (SDOA) is a term that is used in the experimental literature to describe controlled manipulations of visual stimulus presentation relative to participant response. By varying the SDOA in an experimental task, researchers can potentially identify the particular elements of a visual task that cause compatibility effects. The manipulation of SDOA is one that is often used in the study of the Simon Effect and is one that may have special relevance for WIMP-based GUI design. With respect to the sudden onset or drop-out of visual cues such as dialogue boxes or changes in status indication, an evaluation of the Simon Effect may prove to be valuable in understanding how to minimize the impact of unnecessary interruptions when working with a GUI.

In the presented investigation of the UVF and LVF, only the functional advantage of the LVF was characterized. Future work related to the visual fields could look at the perceptual advantages of the UVF, especially with respect to visual tasks that are common to user interaction with graphical displays, such as searching for icons and tracking dynamic (i.e. moving or state-changing) objects. The experimental literature suggests that the functional specialization of the UVF could yield performance improvements similar to the ones observed for pointing in the LVF. Another study could evaluate the UVF and LVF in more ecologically-valid scenarios, especially where awareness of the entire visual field is necessary. It would be interesting to understand what happens when objects of interest in a graphical interface move from being perceived in the LVF to the UVF and vice-versa. The lateral (i.e. left and right) visual fields are believed to have specialized behavioural functions as well, particularly in relation to tasks such as reading and symbolic comprehension. Since these are tasks that are often done on a computer, the functional advantages of the lateral visual fields may be of equal interest to HCI as the vertical visual fields. Characterizing the visual fields laterally and vertically may also be useful in developing a four-quadrant model



---

of human vision (i.e. one quadrant for each combination of the lateral and vertical fields), and how each quadrant of the visual world might be best used in a GUI.

The two-visual systems hypothesis presents several avenues for additional experimentation, particularly with respect to visual cues in a graphical environment and other kinds of visual illusions. In the current experiment, the visual feedback of a pointing interaction was tested, but there may be many other kinds of interactions that could benefit from a two-visual systems model of visual processing. Tangible user interfaces and other systems that lend themselves to true physical, direct manipulation may serve as the basis for future experiments, especially where the primary physical movements involved include reaching or grasping as components of the interaction. Another possible direction for research involving the two-visual systems could involve further characterizing the impact of lag, as in how much lag is necessary before users begin to interpret visual feedback as unreliable. Other experiments could look at the effect of physical versus virtual context cues and the interaction of both, especially those cues that might conflict and cause an implicit perceptual bias in visual judgments made by users. These could be more important as the distinction between physical and virtual interaction continues to blur and there is a greater need to understand how the human visual system might adjust to cope in these kinds of environments.



## Chapter 9

# Conclusion

This dissertation has argued that:

**A representational approach to the study of HCI is able to predict variations in visually-based user performance neglected by existing theoretical frameworks in HCI. Such insight can be helpful in the design and evaluation of interactive systems.**

The importance of a representational approach to HCI is derived from the historical importance of information processing frameworks, such as the Model Human Processor of Card, Moran, and Newell (1983) to the empirical understanding of user performance. In the Model Human Processor framework, mental representations are indicated as a factor influencing the perceptual, cognitive, and motor attributes of the human user although these have only received very limited attention in HCI. Because mental representations have served as a valuable model in cognitive psychology, particularly with respect to computational perspectives on the human mind, it is possible that such representations could be equally valuable to the study of HCI.

The psychological literature suggests that visual information is mentally represented in multiple, simultaneous ways to enable different kinds of visual tasks. Three representational theories related to the study of the relationship between visual perception and action have been chosen to examine how a representational approach to HCI might be used to compliment existing approaches to evaluation in HCI:

1. **Stimulus-response (S-R) compatibility** refers to one way in which visual information is represented as an *integrated* construct during visual processing. The congruency between the position and orientation of a presented visual stimulus and its associated response are believed to influence how efficiently people can configure appropriate motor movements toward objects in the surrounding world. A key assumption, the coding hypothesis, suggests that the recoding between stimulus and response

as internal representations can account for performance differences. The theoretical predictions of S-R compatibility were applied to the GUI design issue of cursors and their directional representations. A controlled experiment comparing mouse, pointer, and pen input demonstrated that when the directional cues of a pointing cursor were congruent with that of movement direction, performance was better than when they were not. Thus, the representational theory of S-R compatibility suggests that user performance can be systematically varied by simply changing the way in which a visual cursor is presented to a user.

2. The functional specialization of the **upper and lower visual fields (UVF/LVF)** refers to one way in which visual information can be viewed as a *separable* construct during visual processing. The UVF and LVF are specialized respectively to draw visual information from separate mental representations for perception and visually-guided movement. The theoretical implications for this functional division were applied to better understand the relationship between perceived spatial location and pointing performance. A controlled experiment comparing mouse and touchscreen input demonstrated that pointing performance was measurably better when pointed items were initially perceived in the LVF. Thus, the representational theory of the specialization of the UVF and LVF suggests that user performance can be systematically varied by simply changing the spatial location at which an interactive element is initially presented to a user.
3. The **two-visual systems hypothesis** refers to one way in which visual information can be viewed as a *mediated* construct during visual processing. Based on a variety of case studies and experimental evidence, the hypothesis states that responses to perceived visual information draw either from a perceptual representation of space encapsulated in a ventral stream of the human brain or from a motor representation of space encapsulated in a dorsal stream of the human brain. The hypothesis further predicts that the dominance of one representation at a given moment can be influenced by the presence or absence of particular visual cues. The implications of the hypothesis were applied to understand how voice input and the presence or absence of visual feedback in large screen pointer interaction compared to one another. A controlled experiment involving the perceptual bias of a visual illusion known as the induced

Roelofs Effect demonstrated that voice input and pointing with a visible cursor were consistently biased by the visual illusion, while pointing without any cursor or even pointing with a lagged cursor were substantially more robust. Thus, the representational theory of the two-visual systems hypothesis suggests that user performance in pointing on a large screen can be systematically varied by simply changing the surrounding visual cues that are present or absent on a graphical display, particularly when users' sense of context is limited to those of the visual cues provided on the display.

The results of these three controlled experiments provide empirical evidence supporting the central claim of this dissertation that representational theories can identify systematic variations in user performance that have remained relatively unstudied by the HCI community. The experiments also demonstrate that improvements to user performance can be had by making simple changes to the visual characteristics that make up the graphical elements of a user interface, suggesting that these results could be valuable in the design and evaluation of future interactive systems. In particular, the results derived from a representational approach to HCI might be beneficial for systems applied to time- and safety-critical situations, interaction with computer graphics, information visualization, and computer-supported cooperative work (CSCW). These are all areas where pointing is an important method for user interaction with a computer, and where understanding visual cognition could be central to providing users with an optimal user experience.

The study of mental representations is likely to continue to be important in HCI, especially with respect to fully understanding mental models and the ways in which users behave when they are placed in a situation where computers are no longer tied to the desktop, but are deployed in immersive environments more like the natural settings in which the human visual and motor systems co-evolved. As technology continues to advance, so must the techniques that are used to understand why and how these technologies can be appropriately used to benefit the tasks that humans perform. The basis that mental representations provide for not only *describing* user performance but also *understanding* the factors that drive performance has only demonstrated a fraction of its usefulness so far. With continued research, an approach driven by the theoretical framework of mental representations may be especially valuable as HCI steadily makes progress toward developing theories about how systems should be designed and evaluated.



## References

- Accot, J., & Zhai, S. (1989). A benchmark comparison of mouse and touch interface techniques for an intelligent workstation windowing environment. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society*.
- Accot, J., & Zhai, S. (1997). Beyond Fitts' law: Models for trajectory-based HCI tasks. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 295–302.
- Accot, J., & Zhai, S. (1999). Performance evaluation of input devices in trajectory-based tasks: An application of the steering law. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 466–472.
- Accot, J., & Zhai, S. (2003). Refining Fitts' law models for bivariate pointing. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 193–200.
- Adler, H. E. (1966). *Elements of Psychophysics*. New York, NY: Holt, Rinehart and Winston.
- Aglioti, S., & DeSouza, J. F. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, 5(6), 679–685.
- Akeley, K., Watt, S. J., Girshick, A. R., & Banks, M. S. (2004). A stereo display prototype with multiple focal distances. *ACM Transactions on Computer Graphics*, 23(3), 804–813.
- Albert, A. F. (1982). The effect of graphic input devices on performance in a cursor positioning task. *Proceedings of the Human Factors Society*, 54–58.
- Allman, J. (1977). Evolution of the visual system in early primates. *Progress in Psychobiology and Physiological Psychology*, 7, 1–53.
- Arthur, K. W., Booth, K. S., & Ware, C. (1993). Evaluating 3D task performance for fish tank virtual worlds. *ACM Transactions on Information Systems*, 11(3), 239–265.
- Austin, J. L. (1962). *Sense and Sensibilia*. New York, NY: Oxford University Press.
- Baecker, R. M., Tilbrook, D. M., Tuori, M. I., & McFarland, D. (1979). NEWSWHOLE. *Proceedings of the International Conference on Computer Graphics and Interactive Techniques, Videotape*.
- Balakrishnan, R., Baudel, T., Kurtenbach, G., & Fitzmaurice, G. (1997). The Rockin'Mouse: Integral 3D manipulation on a plane. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 311–318.

- Baudisch, P., Cutrell, E., Hinckley, K., & Gruen, R. (2004). Mouse ether: Accelerating the acquisition of targets across multi-monitor displays. *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems*, 1379–1382.
- Bauer, D. W., & Miller, J. (1982). Stimulus-response compatibility and the motor system. *The Quarterly Journal of Experimental Psychology*, *34A*, 367–380.
- Biel, G. A., & Carswell, C. M. (1993). Musical notation for the keyboard: An examination of stimulus-response compatibility. *Applied Cognitive Psychology*, *7*, 433–452.
- Bier, E., Stone, M., Pier, K., Buxton, W., & DeRose, T. (1993). Tooglass and magic lenses: The see-through interface. *Proceedings of the ACM Conference on Computer Graphics and Interactive Techniques*, 73–80.
- Bishop, A. (1962). Hand control in lower primates. *Annual New York Academic Science*, *102*, 316–337.
- Bolt, R. A. (1980). “Put-that-there”: Voice and gesture at the graphics interface. *Proceedings of the ACM Conference on Computer Graphics and Interactive Techniques*, 262–270.
- Boritz, J., Booth, K. S., & Cowan, W. B. (1991). Fitts’s law studies of directional mouse movement. *Proceedings of Graphics Interface*, 216–223.
- Brebner, J. (1979). The compatibility of spatial and non-spatial relationships. *Acta Psychologica*, *43*, 23–32.
- Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, *15*, 719–722.
- Bridgeman, B., Kirch, M., & Sperling, A. (1981). Segregation of cognitive and motor aspects of visual function using induced motion. *Perception and Psychophysics*, *29*(4), 336–342.
- Bridgeman, B., Lewis, S., Heit, G., & Nagle, M. (1979). Relation between cognitive and motor-oriented systems of visual position perception. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 692–700.
- Bridgeman, B., Peery, S., & Anand, S. (1997). Interaction of cognitive and sensorimotor maps of visual space. *Perception and Psychophysics*, *59*(3), 456–469.
- Brinck, I. (2004). The pragmatics of imperative and declarative pointing. *Cognitive Science Quarterly*, *3*(4), 1–18.
- Broadbent, D. E. (1971). *Decision and Stress*. London, UK: Academic Press.
- Bush, V. (1945, July). As we may think. *The Atlantic Monthly*.
- Buxton, W. (1986). There’s more to interaction than meets the eye: Some issues in manual input. *User Centered System Design: New Perspectives on Human-Computer Interaction*, 319–337.
- Buyukkokten, O., Garcia-Molina, H., Paepcke, A., & Winograd, T. (2000). Power browser: Efficient Web browsing for PDAs. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 430–437.



- Callahan, J., Hopkins, D., Weiser, M., & Shneiderman, B. (1988). An empirical comparison of pie vs. linear menus. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 95–100.
- Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics*, 601–613.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Cavanagh, P. (2001). Seeing the forest but not the trees. *Nature Neuroscience*, 4, 673–674.
- Cave, K. R., & Kim, M. (1999). Top-down and bottom-up attentional control: On the nature of inference from a salient distractor. *Perception and Psychophysics*, 61, 1009–1023.
- Cavens, D., Vogt, F., Fels, S. S., & Meitner, M. (2002). Interacting with the big screen: Pointers to ponder. *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems*, 678–679.
- Chapanis, A., & Lindenbaum, L. E. (1959). A reaction time study of four control-display linkages. *Human Factors*, 1, 1–7.
- Chedru, F., Leblanc, M., & Chermitte, F. (1973). Visual searching in normal and brain damaged subjects: Contribution to the study of unilateral inattention. *Cortex*, 9, 94–111.
- Cho, Y. S., & Proctor, R. W. (2001). Effect of an initiating action on the up-right/down-left advantage for vertically arrayed stimuli and horizontally arrayed responses. *Journal of Experimental Psychology: Human Perception and Performance*, 27(2), 472–484.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Christian, K., Kules, B., Shneiderman, B., & Youssef, A. (2000). A comparison of voice controlled and mouse controlled web browsing. *Proceedings of the ACM Conference on Assistive Technologies*, 72–79.
- Christman, S. D. (1993). Local-global processing in the upper versus lower visual fields. *Bulletin of the Psychonomic Society*, 31, 275–278.
- Chua, R. (1995). *Informational Constraints in Perception-Action Coupling*. Burnaby, Canada: Simon Fraser University.
- Chua, R., Weeks, D. J., Ricker, K. L., & Poon, P. (2001). Influence of operator orientation on relative organizational mapping and spatial compatibility. *Ergonomics*, 751–765.
- Citrin, W., Halbert, D., Hewitt, C., Meyrowitz, N., & Shneiderman, B. (1993). Potentials and limitations of pen-based computers. *Proceedings of CSC*, 536–539.
- Cohen, S. M., Curd, C., & Reeves, C. D. C. (2000). *Readings in Ancient Greek Philosophy*. Indianapolis, IN: Hackett Publishing Company.
- Craft, J. L., & Simon, J. R. (1970). Processing symbolic information from a visual display: Interference from an irrelevant directional cue. *Journal of Experimental Psychology*, 83, 415–420.

- Cruz-Neira, C., Sandin, D., DeFanti, T., Kenyon, R., & Hart, J. (1992). The CAVE: Audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6), 65–72.
- Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *Journal of Comparative Neurology*, 300(1), 5–25.
- Curcio, C. A., Sloan, K. R., Packer, O., Hendrickson, A. E., & Kalina, R. E. (1987). Distribution of cones in human and monkey retina: Individual variability and radial asymmetry. *Science*, 236, 579–581.
- Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, G., & Starkweather, G. (2003). Toward characterizing the productivity benefits of very large displays. *Proceedings of INTERACT*, 252–259.
- Czerwinski, M., Tan, D. S., & Robertson, G. G. (2002). Women take a wider view. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 195–202.
- Danckert, J., & Goodale, M. A. (2001). Superior performance for visually guided pointing in the lower visual field. *Experimental Brain Research*, 137, 303–308.
- de Bruijn, O., Spence, R., & Chong, M. Y. (2002). RSVP browser: Web browsing on small screen devices. *Personal and Ubiquitous Computing*, 245–252.
- De Jong, R., Liang, C. C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus-response correspondence. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 731–750.
- Dietz, P., & Leigh, D. (2001). DiamondTouch: A multi-user touch technology. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 219–226.
- Dutta, A., & Proctor, R. W. (1992). Persistence of stimulus-response compatibility effects with extended practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 801–809.
- Ehrenstein, W. H., & Ehrenstein, A. (1999). Psychophysical methods. *Modern Techniques in Neuroscience Research*, 1211–1241.
- Ekman, P., & Friesen, W. (1969). The repertoire of non-verbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1), 49–98.
- Engelbart, D. C. (1963). A conceptual framework for the augmentation of man's intellect. *Vistas in Information Handling*, 1.
- English, W. K., Engelbart, D. C., & Berman, M. L. (1967). Display selection techniques for text manipulation. *IEEE Transactions on Human Factors in Electronics*, HFE-8(1), 5–15.
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling amplitude of movement. *Journal of Experimental Psychology*, 47, 381–391.
- Fitts, P. M., & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, 46, 199–210.

- 
- Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Foley, J. D., van Dam, A., Feiner, S. K., & Hughes, J. F. (1997). *Computer Graphics: Principles and Practice*. Boston, MA: Addison-Wesley.
- Frohlich, B., & Plate, J. (2000). The cubic mouse: A new device for three-dimensional input. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 526–531.
- Gawryszewski, L. D. G., Riggio, L., Rizzolatti, G., & Umiltà, C. (1987). Movements of attention in three spatial dimensions and the meaning of neutral cues. *Neuropsychologia*, 25, 19–29.
- Gentilucci, M., Chieffi, S., Daprati, E., Saetti, M. C., & Toni, I. (1996). Visual illusion and action. *Neuropsychologia*, 34(6), 369–376.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton-Mifflin.
- Goldstein, A., & Babkoff, H. (2001). A comparison of upper vs. lower and right vs. left visual fields using lexical decision. *The Quarterly Journal of Experimental Psychology*, 54A(4), 1239–1259.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience*, 15(1), 20–25.
- Goodale, M. A., & Milner, A. D. (2004). *Sight Unseen: An Exploration of Conscious and Unconscious Vision*. New York, NY: Oxford University Press.
- Grudin, J. (2001). Partitioning digital worlds: Focal and peripheral awareness in multiple monitor use. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 458–465.
- Guiard, Y., Beaudouin-Lafon, M., Bastin, J., Pasveer, D., & Zhai, S. (2004). View size and pointing difficulty in multi-scale navigation. *Proceedings of the Working Conference on Advanced Visual Interfaces*, 117–124.
- Guiard, Y., Beaudouin-Lafon, M., & Mottet, D. (1999). Navigation as multiscale pointing: Extending fitts' model to very high precision tasks. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 450–457.
- Gutwin, C., & Fedak, C. (2004). Interacting with big interfaces on small screens: A comparison of fisheye, zooming, and panning techniques. *Proceedings of Graphics Interface*, 145–152.
- Haffenden, A., & Goodale, M. A. (1998). The effect of pictorial illusion on prehension and perception. *Journal of Cognitive Neurosciences*, 10, 122–136.
- Hancock, M. S., & Booth, K. S. (2004). Improving menu placement strategies for pen input. *Proceedings of Graphics Interface*, 221–230.
- Hansen, J. P., Torning, K., Johansen, A. S., Itoh, K., & Aoki, H. (2004). Gaze typing compared with input by head and hand. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 131–138.

- Harrison, J., Rensink, R. A., & van de Panne, M. (2004). Obscuring length changes during animated motion. *ACM Transactions on Graphics, 23*(3), 569–573.
- Heider, B., & Groner, R. (1997). Vertical visual-field differences in schematic persistence for colour information. *Proceedings of the European Conference on Visual Perception*.
- Heywood, S., & Churcher, J. (1980). Structure of the visual array and saccadic latency: Implications for oculomotor control. *Quarterly Journal of Experimental Psychology, 32*, 335–341.
- Hinckley, K. (2003). Synchronous gestures for multiple persons and computers. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 149–158.
- Hinckley, K., Cutrell, E., Bathiche, S., & Muss, T. (2002). Quantitative analysis of scrolling techniques. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 65–72.
- Hinckley, K., Pausch, R., Goble, J. C., & Kassell, N. F. (1994a). Passive real-world interface props for neurosurgical visualization. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 452–458.
- Hinckley, K., Pausch, R., Goble, J. C., & Kassell, N. F. (1994b). A survey of design issues in spatial input. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 213–222.
- Hinckley, K., Sinclair, M., Hanson, E., Szeliski, R., & Conway, M. (1999). The VideoMouse: A camera-based multi-degree-of-freedom input device. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 103–112.
- Hinckley, K., Tullio, J., Pausch, R., Proffitt, D., & Kassell, N. (1997). Usability analysis of 3D rotation techniques. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 1–10.
- Hommel, B. (1995). Conflict versus misguided search as explanation of S-R correspondence effects. *Acta Psychologica, 89*, 37–51.
- Hutchings, D. R., & Stasko, J. (2004). Shrinking window operations for expanding display space. *Proceedings of the Working Conference on Advanced Visual Interfaces*, 350–353.
- Igarashi, T., & Hughes, J. F. (2001). Voice as sound: Using non-verbal voice input for interactive control. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 155–156.
- Inkpen, K. M. (2001). Drag-and-drop versus point-and-click mouse interaction styles for children. *ACM Transactions on Computer-Human Interaction, 8*(1), 1–33.
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*, 171–216.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506.

- 
- Izadi, S., Brignull, H., Rodden, T., Rogers, Y., & Underwood, M. (2003). Dynamo: A public surface supporting the cooperative sharing and exchange of media. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 159–168.
- Jackson, S. R. (2000). Perception, awareness, and action: Insights from blindsight. *Beyond Dissociation: Interaction Between Dissociated Implicit and Explicit Processing*, 73–98.
- Jacob, R. J. K., Ishii, H., Pangaro, G., & Patten, J. (2002). A tangible interface for organizing information using a grid. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 339–346.
- Jacob, R. J. K., Sibert, L. E., McFarlane, D. C., & Mullen, M. P. (1994). Integrality and separability of input devices. *ACM Transactions on Computer-Human Interaction*, 1(1), 3–26.
- Jeannerod, M. (1986). The formation of finger grip during prehension: A cortically mediated visuomotor pattern. *Behavioural Brain Research*, 19, 99–116.
- Jeannerod, M., Gerin, P., & Pernier, J. (1968). Deplacements et fixations du regard dans l'exploration d'une scene visuelle. *Vision Research*, 8, 81–97.
- John, B. E., & Newell, A. (1989). Cumulating the science of HCI: From S-R compatibility to transcription typing. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 109–114.
- John, B. E., Rosenbloom, P. S., & Newell, A. (1985). A theory of stimulus-response compatibility applied to human-computer interaction. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 213–219.
- Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge, UK: Cambridge University Press.
- Kantowitz, B. H., Triggs, T. J., & Barnes, V. (1990). Stimulus-response compatibility and human factors. *Stimulus-Response Compatibility*, 365–388.
- Kasik, D. J., Troy, J. J., Amorosi, S. R., Murray, M. O., & Swamy, S. N. (2002). Evaluating graphic displays for complex 3D models. *IEEE Computer Graphics and Applications*, 22(3), 56–64.
- Kawash, J. (2004). Declarative user interfaces for handheld devices. *Proceedings of the ACM Winter International Symposium on Information and Communication Technologies*, 1–6.
- Kay, A. (1977, September). Microelectronics and the personal computer. *Scientific American*, 230–244.
- Kendon, A. (1996). An agenda for gesture studies. *The Semiotic Review of Books*, 7(3), 8–12.
- Kerzel, D., Hommel, B., & Bekkering, H. (2001). A Simon effect induced by induced motion: Evidence for a direct linkage between cognitive and motor maps. *Perception and Psychophysics*, 63(5), 862–874.

- Khan, A., Fitzmaurice, G., Almeida, D., Burtnyk, N., & Kurtenbach, G. (2004). A remote control interface for large displays. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 127–136.
- Kinsbourne, M. (1972). Eye and head turning indicates cerebral lateralization. *Science*, *176*, 539–541.
- Kirstein, A., & Mueller, H. (1998). Interaction with a projection screen using a camera-tracked laser pointer. *Proceedings of the International Conference on Multimedia Modeling*.
- Kornblum, S., Hasbroucq, T., & Osman, A. (1980). Dimensional overlap: Cognitive basis for stimulus-response compatibility - A model and taxonomy. *Psychological Review*, *97*, 253–270.
- Kosslyn, S. M. (1987). Seeing and imaging in the cerebral hemispheres: A computational approach. *Psychological Review*, *94*, 148–175.
- Liang, J., & Green, M. (1993). JDCAD: A highly interactive 3D modeling system. *Proceedings of the 3rd International Conference on CAD and Computer Graphics*, 217–222.
- Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics*, *HFE-1*, 4–11.
- Lippa, Y. (1996). A referential-coding explanation for compatibility effects of physically orthogonal stimulus and response dimensions. *The Quarterly Journal of Experimental Psychology*, *49A*, 950–971.
- Lippa, Y., & Adam, J. (2001). An explanation of orthogonal S-R compatibility effects that vary with hand or response position. *Perception and Psychophysics*, *63*, 156–174.
- Long, G. M., & Toppino, T. C. (1994). Adaptation effects and reversible figures: A comment on Horlitz and O’Leary. *Perception and Psychophysics*, *56*, 605–610.
- MacKenzie, I. S. (1989). A note on the information-theoretic basis for Fitts’ law. *Journal of Motor Behavior*, *21*, 323–330.
- MacKenzie, I. S., & Buxton, W. (1992). Extending Fitts’ law to two-dimensional tasks. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 219–226.
- MacKenzie, I. S., Sellen, A., & Buxton, W. (1991). A comparison of input devices in elemental pointing and dragging tasks. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 161–166.
- Mackinlay, J. D., Card, S. K., & Robertson, G. G. (1990). A semantic analysis of the design space of input devices. *Human Computer Interaction*, *5*, 145–190.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman and Company.
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, Series B*, *200*, 269–294.

- 
- Mateef, S., & Gourevich, A. (1983). Peripheral vision and perceived visual direction. *Biological Cybernetics*, *49*, 111–118.
- Michaels, C. F. (1988). S-R compatibility between response position and destination of apparent motion: Evidence of the detection of affordances. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 231–240.
- Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1994). Augmented reality: A class of displays on the reality-virtuality continuum. *Proceedings of SPIE Telemanipulator and Telepresence Technologies*.
- Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 81–97.
- Miller, G. A. (2004). MS-Taxonomy: A conceptual framework for designing multi-sensory displays. *Proceedings of the 8th International Conference on Information Visualisation*, 665–670.
- Milner, A. D., & Goodale, M. A. (1995). *The Visual Brain in Action*, Oxford Psychology Series (Vol. 27). New York, NY: Oxford University Press.
- Morrin, R. E., & Forrin, B. (1962). Mixing two types of S-R associations in a choice reaction time task. *Journal of Experimental Psychology*, *64*, 137–141.
- Myers, B. A., Bhatnagar, R., Nichols, J., Peck, C. H., Kong, D., Miller, R., et al. (2002). Interacting at a Distance: Measuring the performance of laser pointers and other devices. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 33–40.
- Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Niebauer, C. L., & Christman, S. D. (1998). Upper and lower visual field differences in categorical and coordinate judgments. *Psychonomic Bulletin and Review*, *5*, 147–151.
- Nilsson, D.-E. (1996). Eye ancestry: Old genes for new eyes. *Current Biology*, *6*(1), 39–42.
- Oh, J.-Y., & Stuerzlinger, W. (2002). Laser pointers as collaborative pointing devices. *Proceedings of Graphics Interface*, 141–149.
- Olsen, D. R., & Nielsen, T. (2001). Laser pointer interaction. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 17–22.
- Osman Hill, W. C. (1972). *Evolutionary Biology of the Primates*. New York, NY: Academic Press.
- O’Sullivan, C., & Dingliana, J. (2001). Collisions and perception. *ACM Transactions on Graphics*, *20*(3), 151–168.
- O’Sullivan, C., Dingliana, J., Giang, T., & Kaiser, M. K. (2003). Evaluating the visual fidelity of physically based animations. *Proceedings of the ACM Conference on Computer Graphics and Interactive Techniques*, 527–536.
- Oviatt, S. L., & Cohen, P. R. (2000). Multimodal interfaces that process what comes naturally. *Communications of the ACM*, *43*(3), 45–53.

- Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. New York, NY: Oxford University Press.
- Parker, J. K., Mandryk, R. L., & Inkpen, K. M. (2004). TractorBeam: Seamless integration of local and remote pointing for tabletop displays. *Technical Report CS-2004-09*.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. J. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4, 739–744.
- Pausch, R., Proffitt, D., & Williams, G. (1997). Quantifying immersion in virtual reality. *Proceedings of the ACM Conference on Computer Graphics and Interactive Techniques*, 13–18.
- Payne, S. J. (1995). Naive judgments of stimulus-response compatibility. *Human Factors*, 37, 495–506.
- Payne, W. H. (1967). Visual reaction time on a circle about the fovea. *Science*, 155, 481–482.
- Pelisson, D., Prablanc, C., Goodale, M. A., & Jeannerod, M. (1986). Visual control of reaching movements without vision of the limb. *Experimental Brain Research*, 62, 303–311.
- Perenin, M. T., & Rossetti, Y. (1996). Grasping in a hemianopic field: Another instance of dissociation between perception and action. *Neuroreport*, 7(3), 793–797.
- Perenin, M. T., & Vighetto, A. (1988). Optic ataxia: A specific disruption in visuomotor mechanisms. *Brain*, 111(3), 643–674.
- Pham, T.-L., Schneider, G., & Goose, S. (2000). A situated computing framework for mobile and ubiquitous multimedia access using small screen and composite devices. *Proceedings of the ACM Conference on Multimedia*, 323–331.
- Phillips, J. G., Meehan, J. W., & Triggs, T. J. (2003). Effects of cursor orientation and required precision on positioning movements on computer screens. *International Journal of Human-Computer Interaction*, 15(3), 379–389.
- Phillips, J. G., Triggs, T. J., & Meehan, J. W. (2003). Conflicting directional and locational cues afforded by arrow cursors in graphical user interfaces. *Journal of Experimental Psychology: Applied*, 9(2), 75–87.
- Piaget, J. (1969). *The Mechanisms of Perception*. London, UK: Rutledge and Kegan Paul.
- Picard, R. W. (1997). *Affective Computing*. Cambridge, MA: MIT Press.
- Pinker, S., & Mehler, J. (1988). *Connections and Symbols*. Cambridge, MA: MIT Press.
- Pisella, L., & Rossetti, Y. (2000). Interaction between conscious identification and non-conscious sensorimotor processing: Temporal constraints. *Beyond Dissociation: Interaction Between Dissociated Implicit and Explicit Processing*, 129–151.
- Polyak, S. (1957). *The Vertebrate Visual System*. Chicago: University of Chicago Press.
- Potter, R. L., Weldon, L. J., & Shneiderman, B. (1988). Improving the accuracy of touch screens: An experimental evaluation of three strategies. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 27–32.



- 
- Poupyrev, I., Billingham, M., Weghorst, S., & Ichikawa, T. (1996). Go-Go interaction technique: Non-linear mapping for direct manipulation in VR. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 79–80.
- Poupyrev, I., Billingham, M., Weghorst, S., & Ichikawa, T. (1998). Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques. *Computer Graphics Forum, Eurographics*, 41–52.
- Previc, F. H. (1990). Functional specialization in the lower and upper visual fields in humans: Its ecological origins and neurophysiological implications. *Behavioral and Brain Sciences*, 13, 519–575.
- Price, L. A. (1984). Studying the mouse for CAD systems. *Proceedings of the Conference on Design Automation*, 288–293.
- Prinz, W., & Hommel, B. (2002). *Common Mechanisms in Perception and Action: Attention and Performance XIX*. New York, NY: Oxford University Press.
- Proctor, R. W., Dutta, A., Kelly, P. L., & Weeks, D. J. (1994). Cross-modal compatibility effects with visuo-spatial and audio-verbal stimulus and response sets. *Perception and Psychophysics*, 35, 42–47.
- Proctor, R. W., & Reeve, T. G. (1990). *Stimulus-Response Compatibility An Integrated Perspective*. Amsterdam: North-Holland.
- Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. W. (2003). *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. W., Burkell, J., Fisher, B., Sears, C., Schmidt, W., & Trick, L. (1994). Multiple parallel access in visual attention. *Canadian Journal of Experimental Psychology*, 48(2), 260–283.
- Raskar, R., Brown, M. S., Yang, R., Chen, W.-C., Welch, G., Towles, H., et al. (1999). Multi-projector displays using camera-based registration. *Proceedings of the Conference on Visualization*, 161–168.
- Ren, X., & Moriya, S. (2000). Improving selection performance on pen-based systems: A study of pen-based interaction for selection tasks. *ACM Transactions on Computer-Human Interaction*, 7(3), 384–416.
- Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision Research*, 40, 1469–1487.
- Rensink, R. A. (2002a). Change detection. *Annual Review of Psychology*, 53, 245–277.
- Rensink, R. A. (2002b). Internal vs. external information in visual perception. *Proceedings of the Second International Symposium on Smart Graphics*, 63–70.
- Ringel, M. (2003). When one isn't enough: An analysis of virtual desktop usage strategies and their implications for design. *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems*, 762–763.

- Ringel, M., Ryall, K., Shen, C., Forlines, C., & Vernier, F. (2004). Release, relocate, reorient, resize: Fluid techniques for document sharing on multi-user interactive tables. *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems*, 1441–1444.
- Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin and Review*, 9, 507–513.
- Rizzolatti, G., Gentilucci, M., & Matelli, M. (1985). Selective spatial attention: One center, one circuit, or many circuits? *Attention and Performance XI*, 251–265.
- Rizzolatti, G., Riggio, L., Dascola, I., & Umiltà, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, 25, 30–36.
- Robbins, D. C., Cutrell, E., Sarin, R., & Horvitz, E. (2004). ZoneZoom: Map navigation for smartphones with recursive view segmentation. *Proceedings of the Working Conference on Advanced Visual Interfaces*, 231–234.
- Roelofs, C. (1935). Optische localisation. *Archiv fur Augenheilkunde*, 109, 395–415.
- Rubens, A. B. (1985). Caloric stimulation and unilateral visual neglect. *Neurology*, 35, 1019–1024.
- Rubin, N., Nakayama, K., & Shapley, R. (1996). Enhanced perception of illusory contours in the lower versus upper visual hemifields. *Science*, 271, 651–653.
- Sarkar, M., Snibbe, S. S., Tversky, O. J., & Reiss, S. P. (1993). Stretching the rubber sheet: A metaphor for viewing large layouts on small screens. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 81–91.
- Sasse, M. A. (1997). *Eliciting and Describing Users' Models of Computer Systems*, Ph.D. Thesis. Birmingham, England: The University of Birmingham.
- Schneider, G. E. (1969). Two visual systems: Brain mechanisms for localization and discrimination are dissociated by tectal and cortical lesions. *Science*, 163, 895–902.
- Schon, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York, NY: Basic Books.
- Searle, J. (1997). *The Mystery of Consciousness*. New York, NY: New York Review Press.
- Sears, A., & Shneiderman, B. (1991). High precision touchscreens: Design strategies and comparisons with a mouse. *International Journal of Man-Machine Studies*, 43(4), 593–613.
- Seetzen, H., Heidrich, W., Stuerzlinger, W., Ward, G., Whitehead, L., Trentacoste, M., et al. (2004). High dynamic range display systems. *ACM Transactions on Computer Graphics*, 23(3), 760–768.
- Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communications*. Urbana, IL: University of Illinois Press.

- 
- Shen, C., Vernier, F. D., Forlines, C., & Ringel, M. (2004). DiamondSpin: An extensible toolkit for around-the-table interaction. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 167–174.
- Shneiderman, B. (1982). The future of interactive systems and the emergence of direct manipulation. *Behaviour and Information Technology*, 1, 237–256.
- Shneiderman, B. (1983). Direct manipulation: A step beyond programming languages. *IEEE Computer*, 16, 57–69.
- Shneiderman, B. (1992). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Boston, MA: Addison-Wesley.
- SIGCHI. (1992). *ACM SIGCHI Curricula for HCI*. New York, NY: ACM Press.
- SIGGRAPH. (1977). *Specifications on the SIGGRAPH/GSPC Core System*. New York, NY: ACM Press.
- Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, 81, 174–176.
- Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, 51, 300–304.
- Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Science*, 1, 261–267.
- Skinner, B. F. (1953). *Science and Human Behavior*. New York, NY: Macmillan.
- Skrandies, W. (1987). The upper and lower visual field of man: Electrophysiological and functional differences. *Progress in Sensory Physiology* 8, 162–189.
- Smith, D. (1977). *Pygmalion: A Computer Program to Model and Stimulate Creative Thought*. Stuttgart, Basel: Birkhauser Verlag.
- Smith, P., & Brebner, J. (1983). S-R compatibility: The relative effects of “relevant” spatial and non-spatial variables. *Australian Journal of Psychology*, 35, 1–10.
- Snodderly, D. M. (1979). Visual discrimination encountered in food foraging by a neotropical primate: Implications for the evolution of color vision. *Behavioral Significance of Color Vision*, 238–279.
- Stewart, J., Bederson, B. B., & Druin, A. (1999). Single display groupware: A model for co-present collaboration. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 286–293.
- Stoffer, T. H. (1991). Attentional focusing and spatial stimulus-response compatibility. *Psychological Research*, 53, 127–135.
- Sturmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures and postures: A study on imitation. *Journal of Experimental Psychology: Human Perception and Performance*, 26(6), 1746–1759.

- Sutherland, I. E. (1963). SketchPad: A man-machine graphical communication system. *AFIPS Conference Proceedings*, 23, 323–328.
- Sutherland, I. E. (1965). The ultimate display. *IFIPS Conference Proceedings*, 2, 506–508.
- Swindells, C., Po, B. A., Hajshirmohammadi, I., Corrie, B., Dill, J., Fisher, B. D., et al. (2004). Comparing CAVE, wall, and desktop displays for navigation and wayfinding in complex 3D models. *Proceedings of Computer Graphics International*, 420–427.
- Tan, D., & Czerwinski, M. (2003). Information voyeurism: Social impact of physically large displays on information privacy. *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems*, 748–749.
- Thagard, P. (1996). *Mind: Introduction to Cognitive Science*. Cambridge, MA: MIT Press.
- Tilbrook, D. M. (1976). *A Newspaper Pagination System, M.Sc. Thesis*. Toronto, ON: The University of Toronto.
- Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. *Psychonomic Bulletin and Review*, 9, 314–318.
- Tlauka, M. (2004). Display-control compatibility: The relationship between performance and judgments of performance. *Ergonomics*, 47(3), 281–295.
- Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologische Forschung*, 31, 299–337.
- Tse, E., & Greenberg, S. (2004). Rapidly prototyping Single Display Groupware through the SDGToolkit. *Proceedings of the 5th Conference on Australasian User Interfaces*, 28, 101–110.
- Tychsen, L., & Lisberger, S. G. (1986). Visual motion processing for the initiation of smooth pursuit eye movements in humans. *Journal of Neurophysiology*, 56, 953–968.
- Ullmer, B., & Ishii, H. (2002). Emerging frameworks for tangible user interfaces. *IBM Systems Journal*, 393(3), 915–931.
- Ungerleider, L. G., & Mishkin, M. (1982). *Analysis of Visual Behaviour*. Cambridge, MA: MIT Press.
- Vicente, K. J., & Torenvliet, G. L. (2000). The Earth is spherical ( $p < 0.05$ ): Alternative methods of statistical inference. *Theoretical Issues in Ergonomic Science*, 1(3), 248–271.
- Vogt, F., Wong, J., Po, B. A., Argue, R., Fels, S. S., & Booth, K. S. (2004). Exploring collaboration with group pointer interaction. *Proceedings of Computer Graphics International*, 636–639.
- Vu, K.-P. L., & Proctor, R. W. (2003). Naive and experienced judgments of stimulus-response compatibility: Implications for interface design. *Ergonomics*, 46(1-3), 169–187.
- Waldrop, M. M. (2001). *The Dream Machine: J. C. R. Licklider and the Revolution that Made Computing Personal*. New York, NY: Penguin.

- 
- Want, R., Schilit, B. N., Adams, N. I., Gold, R., Petersen, K., Goldberg, D., et al. (1995, December). An overview of the ParcTab ubiquitous computing experiment. *IEEE Personal Communications*, 28–43.
- Ware, C. (2004). *Information Visualization: Perception for Design, Second Edition*. San Francisco, CA: Morgan Kaufmann.
- Ware, C., Arthur, K., & Booth, K. S. (1993). Fish tank virtual reality. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 37–42.
- Ware, C., & Balakrishnan, R. (1994). Reaching for objects in VR displays: Lag and frame rate. *ACM Transactions on Computer-Human Interaction*, 1(4), 331–356.
- Ware, C., & Osborne, S. (1990). Exploration and virtual camera control in virtual three dimensional environments. *Proceedings of the ACM Symposium on Interactive 3D Graphics*, 175–183.
- Ware, C., & Rose, J. (1999). Rotating virtual objects with real handles. *ACM Transactions on Computer-Human Interaction*, 6(2), 162–180.
- Watson, J. (1913). Psychology as a behaviorist views it. *Psychological Review*, 20, 158–177.
- Weeks, D. J., Chua, R., & Hamblin, K. (1996). Attention shifts and the Simon effect: A failure to replicate Stoffer (1991). *Psychological Research*, 58, 246–253.
- Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications*. New York, NY: Oxford University Press.
- Welford, A. T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years progress. *Ergonomics*, 3, 189–230.
- Wist, E. R., Ehrenstein, W. H., & Schrauf, M. (1998). A computer-assisted test for the electrophysiological and psychophysical measurement of dynamic visual function based on motion contrast. *Journal of Neuroscience Methods*, 80, 41–47.
- Wolf, C. (1865). Recherches sur l'équation personnelle dans les observations de passages, sa détermination absolue, ses lois, et son origine. *Comptes rendus des séances de l'academie des sciences*, 60, 1268–1272.
- Woodworth, R. S. (1938). *Experimental Psychology*. New York, NY: Holt.
- Worringham, C. J., & Beringer, D. B. (1989). Operator orientation and compatibility in visual-motor task performance. *Ergonomics*, 32, 387–399.
- Worringham, C. J., & Beringer, D. B. (1998). Directional stimulus-response compatibility: A test of three alternative principles. *Ergonomics*, 41(6), 864–880.
- Wynn, T. (2002). Archaeology and cognitive evolution. *Behavioral and Brain Sciences*, 25, 389–438.
- Xerox. (1980). *Xerox Document System Reference Manual*. Palo Alto, CA: Xerox Office Products Division.

Yee, K.-P. (2003). Peephole displays: Pen interaction on spatially aware handheld computers. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1–8.

Zhai, S., Sue, A., & Accot, J. (2002). Movement model, hits distribution and learning in virtual keyboarding. *ACM Conference on Human Factors in Computing Systems*, 17–24.

## Appendix A

# Supplemental Data for Experiment

## 1

The following are additional data, which may be helpful in a supplemental analysis of Experiment 1, as presented in Chapter 5.

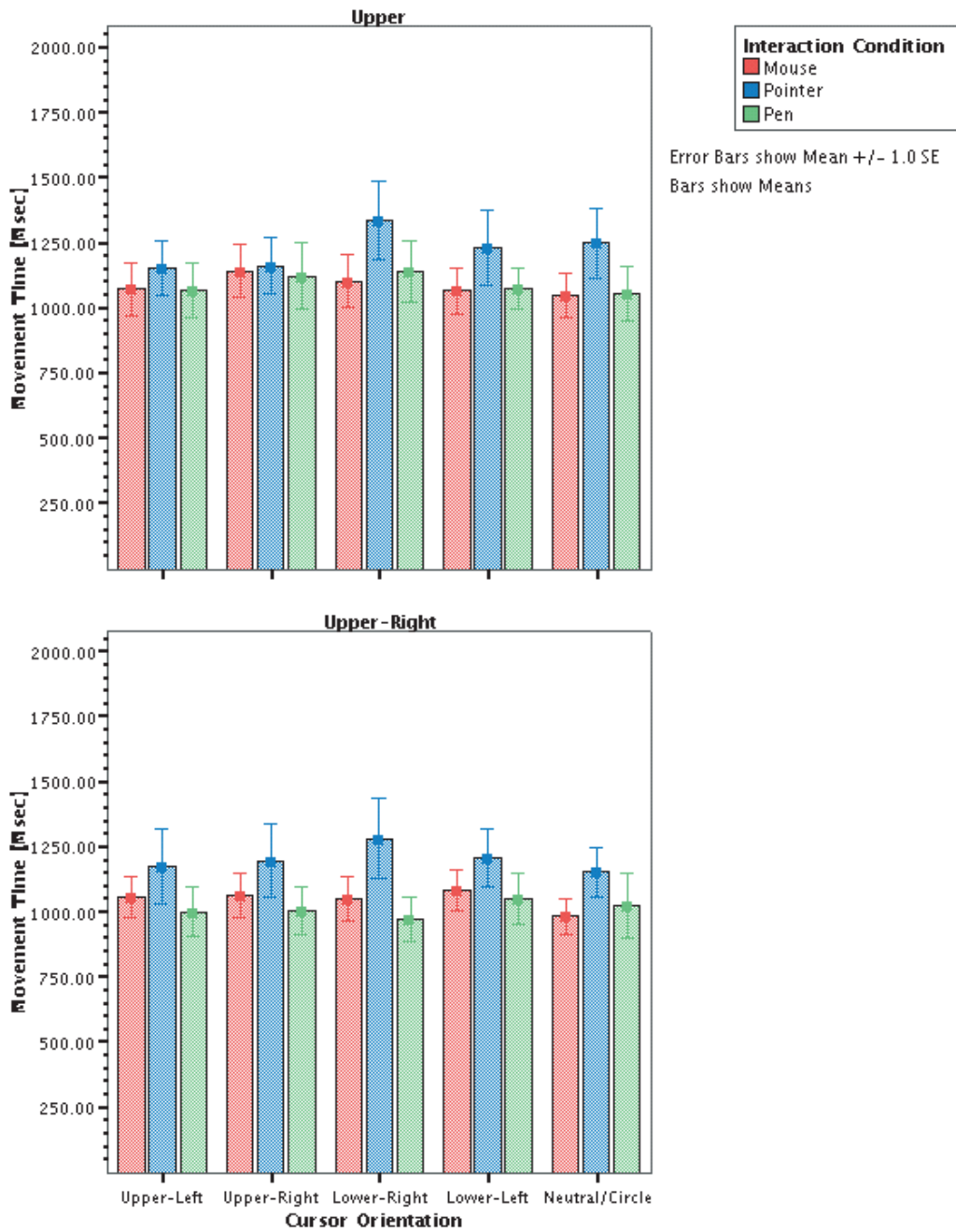


Figure A.1: These bar graphs plot measured movement time across cursor orientations for each input device. Each bar graph represents one menu position, as indicated at the top of each graph.



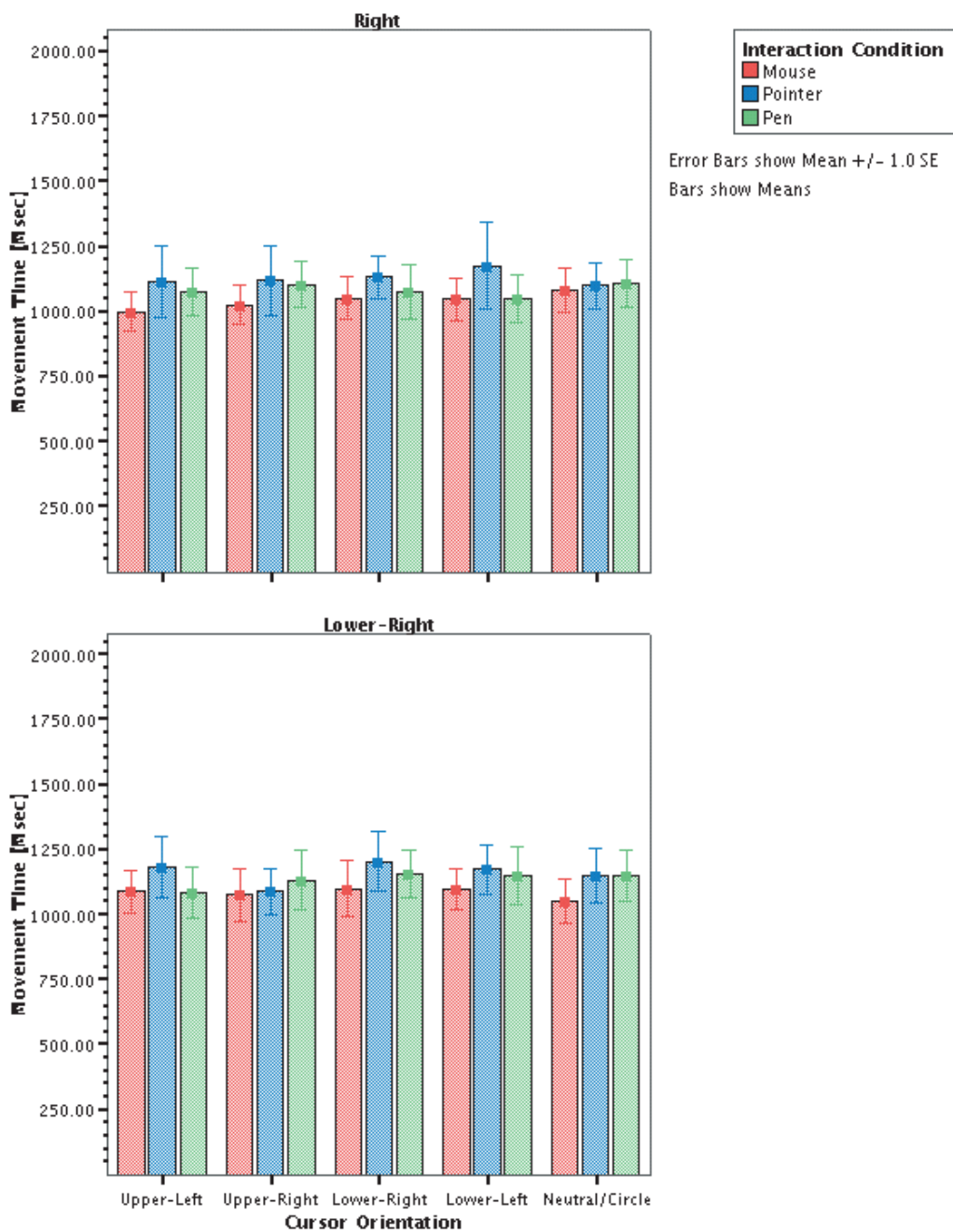


Figure A.2: These bar graphs plot measured movement time across cursor orientations for each input device. Each bar graph represents one menu position, as indicated at the top of each graph.

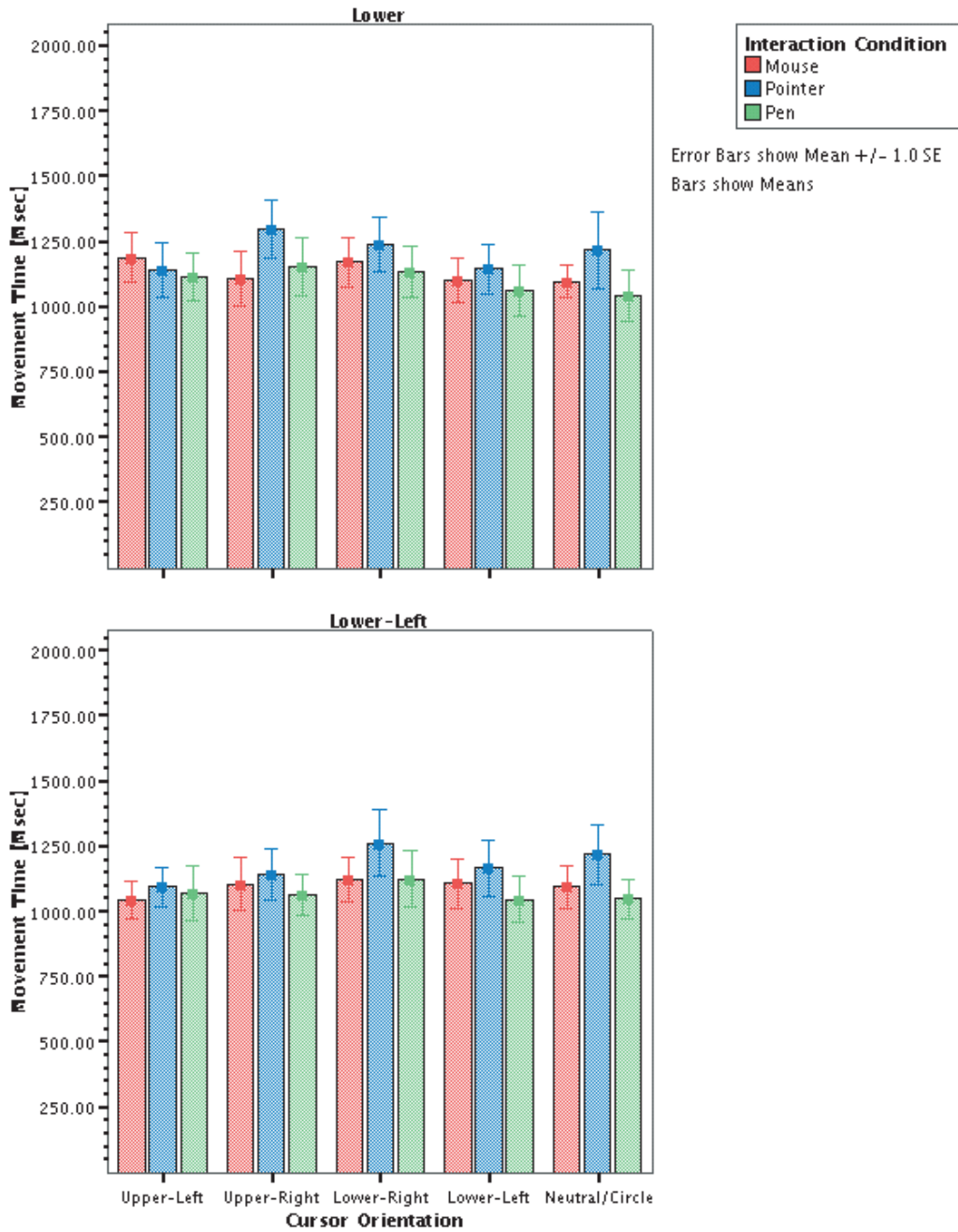


Figure A.3: These bar graphs plot measured movement time across cursor orientations for each input device. Each bar graph represents one menu position, as indicated at the top of each graph.

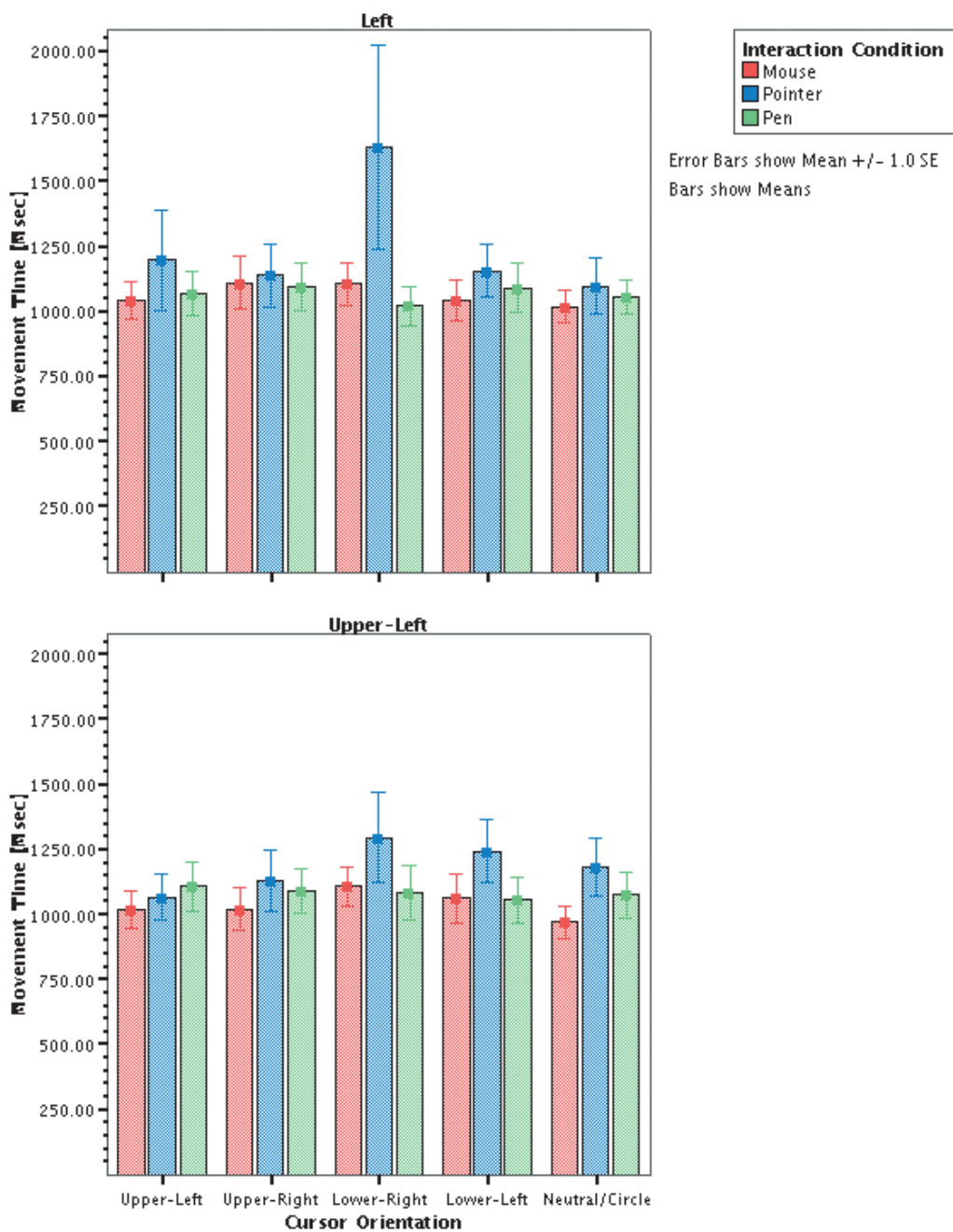


Figure A.4: These bar graphs plot measured movement time across cursor orientations for each input device. Each bar graph represents one menu position, as indicated at the top of each graph.

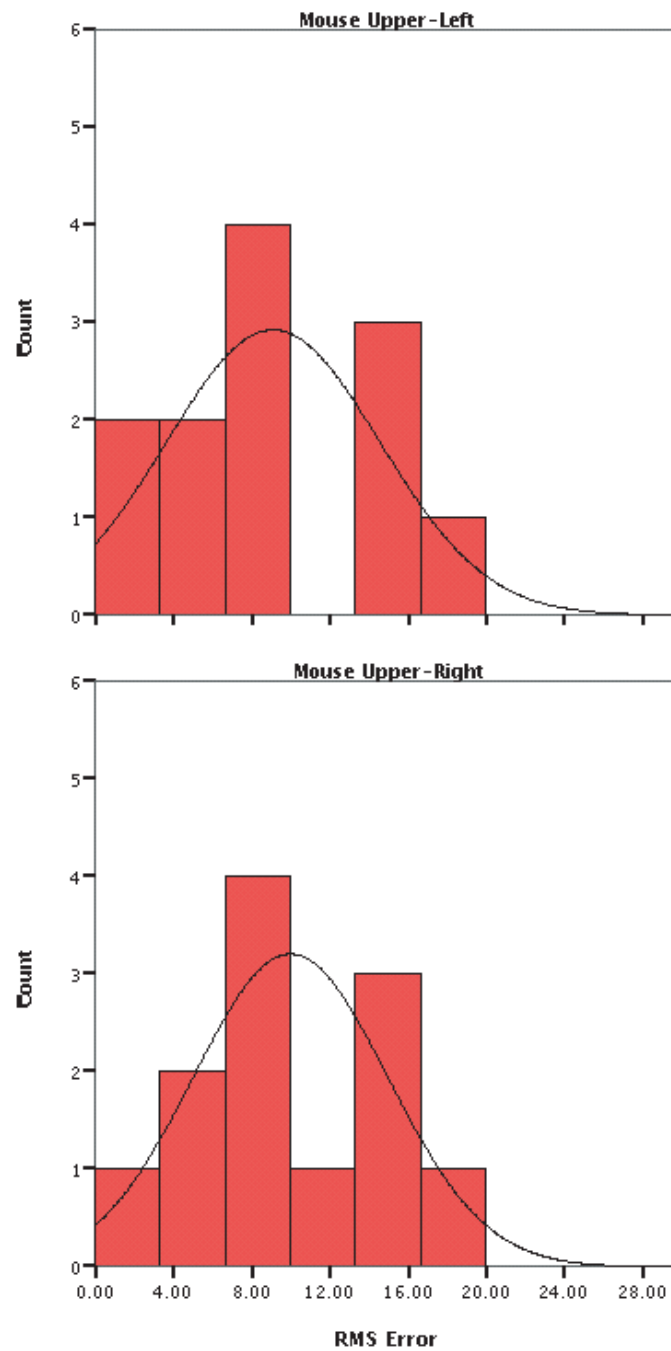


Figure A.5: These histograms present the distribution of RMS errors for each input device and cursor orientation. Each histogram illustrates the distribution of a single cursor orientation and a single input device. The plotted curve indicates the best-fitting normal curve for this particular combination of conditions.

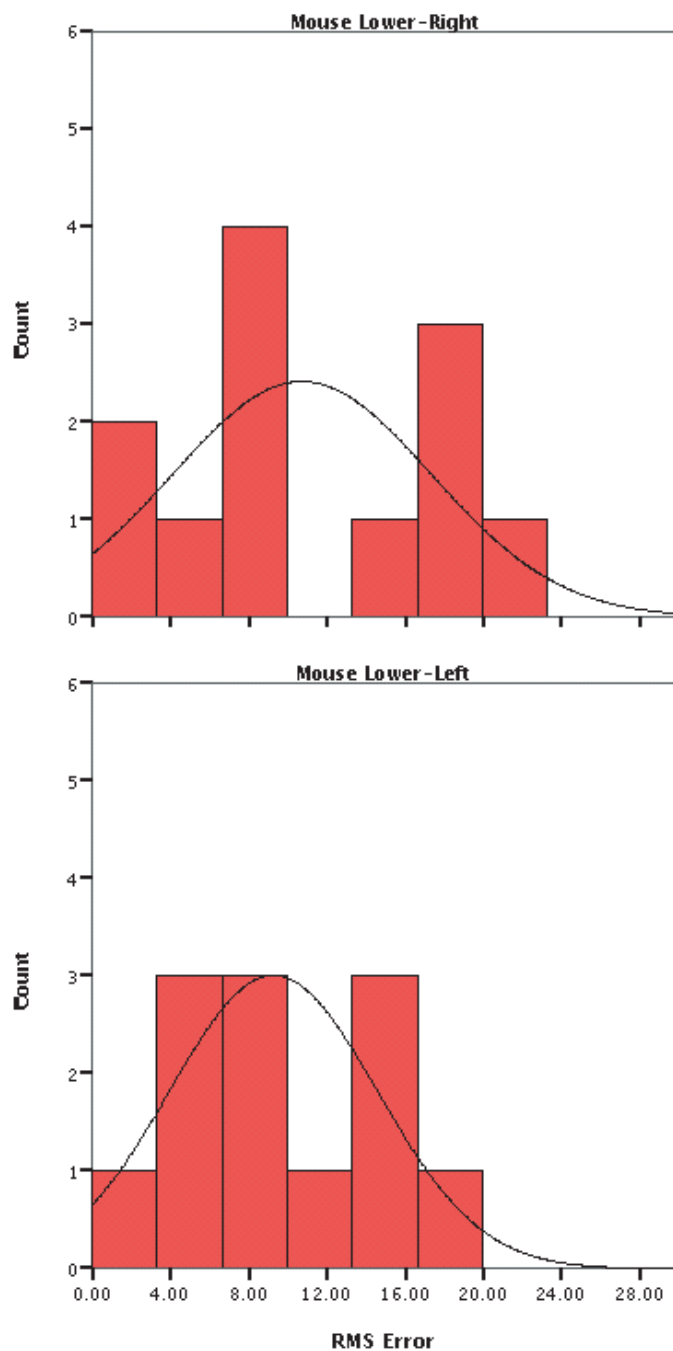


Figure A.6: These histograms present the distribution of RMS errors for each input device and cursor orientation. Each histogram illustrates the distribution of a single cursor orientation and a single input device. The plotted curve indicates the best-fitting normal curve for this particular combination of conditions.

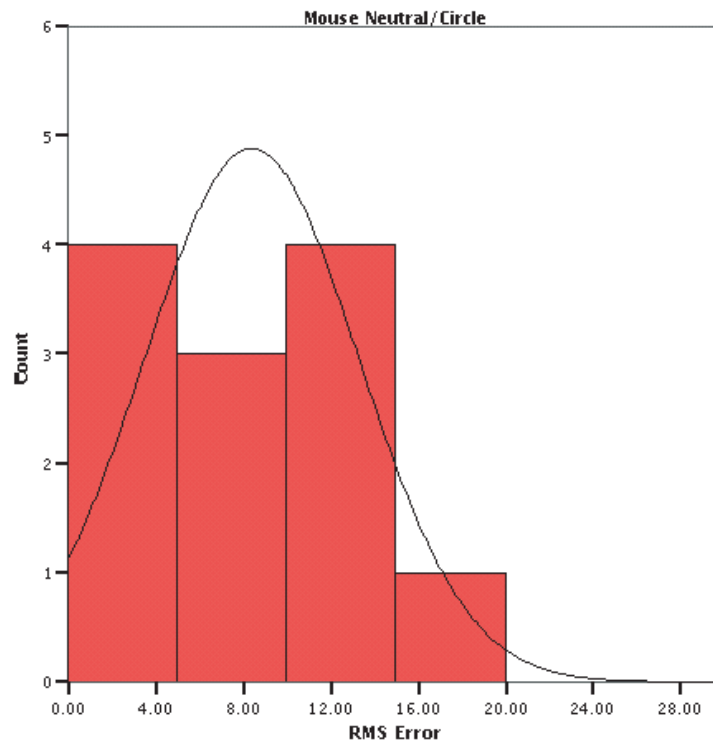


Figure A.7: These histograms present the distribution of RMS errors for each input device and cursor orientation. Each histogram illustrates the distribution of a single cursor orientation and a single input device. The plotted curve indicates the best-fitting normal curve for this particular combination of conditions.

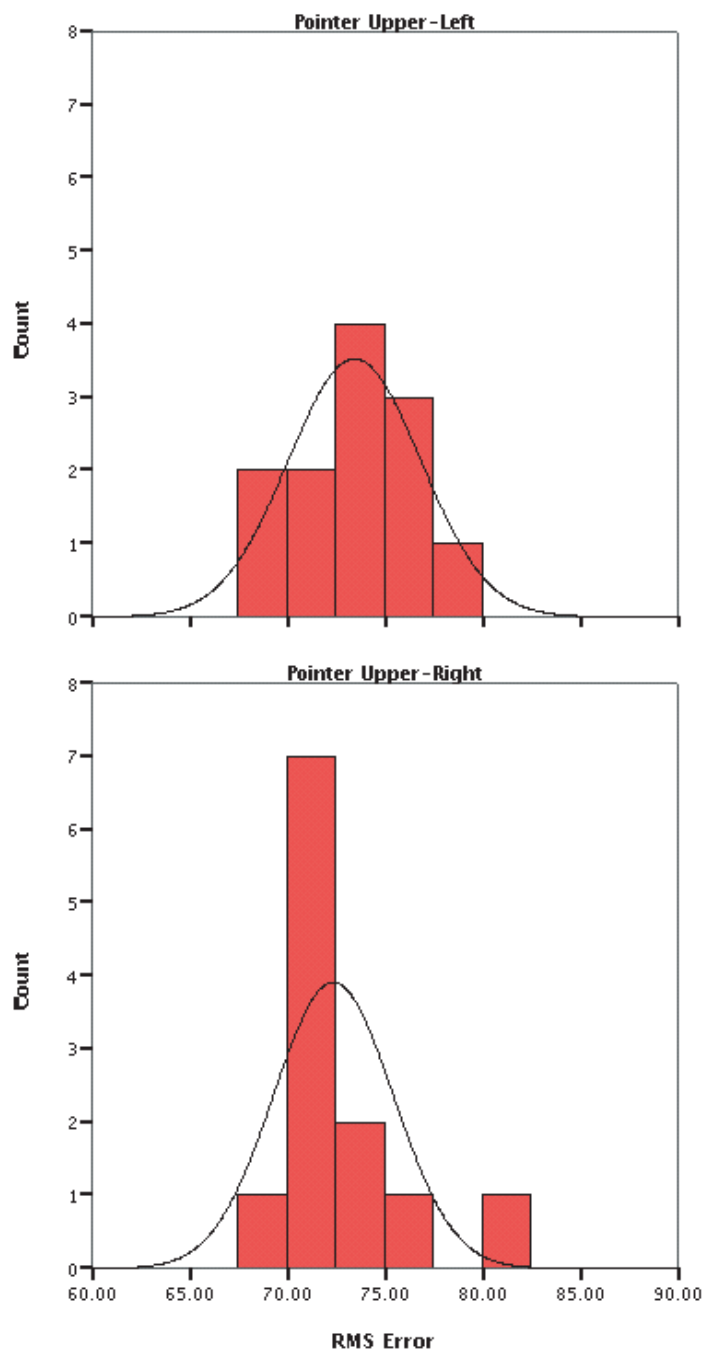


Figure A.8: These histograms present the distribution of RMS errors for each input device and cursor orientation. Each histogram illustrates the distribution of a single cursor orientation and a single input device. The plotted curve indicates the best-fitting normal curve for this particular combination of conditions.

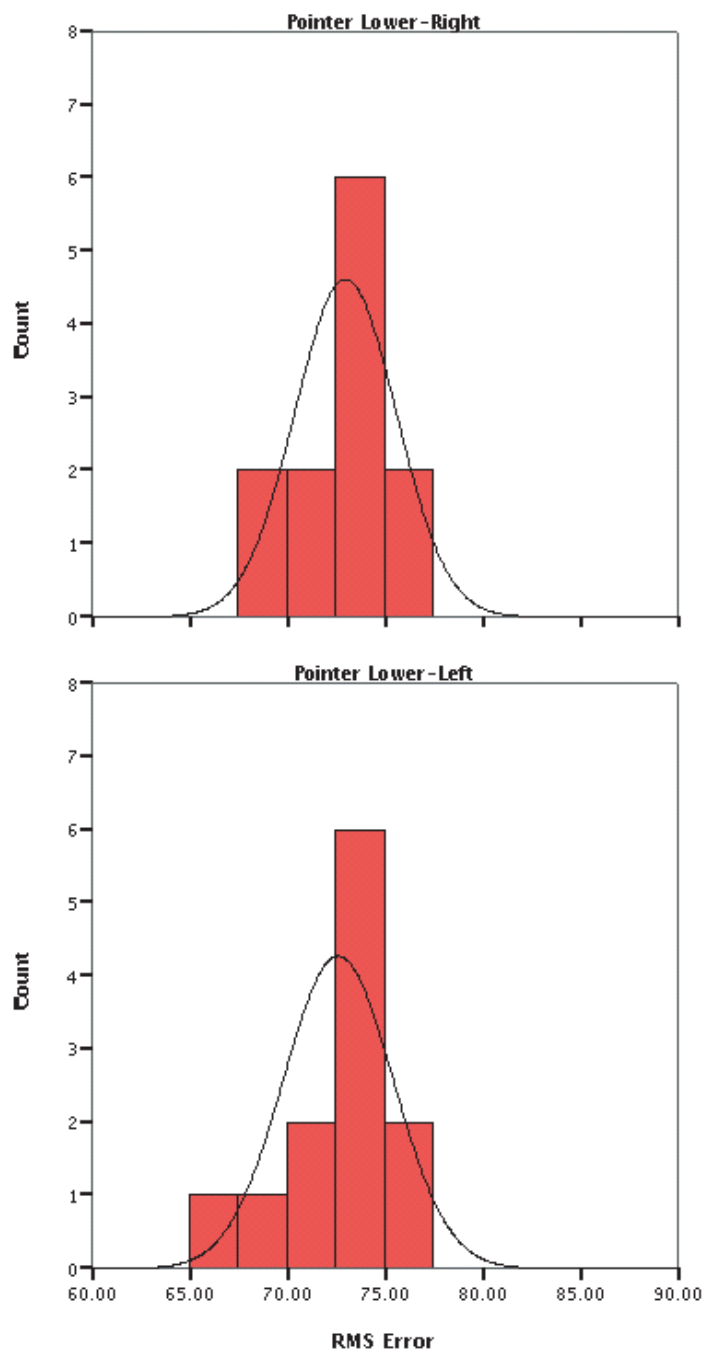


Figure A.9: These histograms present the distribution of RMS errors for each input device and cursor orientation. Each histogram illustrates the distribution of a single cursor orientation and a single input device. The plotted curve indicates the best-fitting normal curve for this particular combination of conditions.



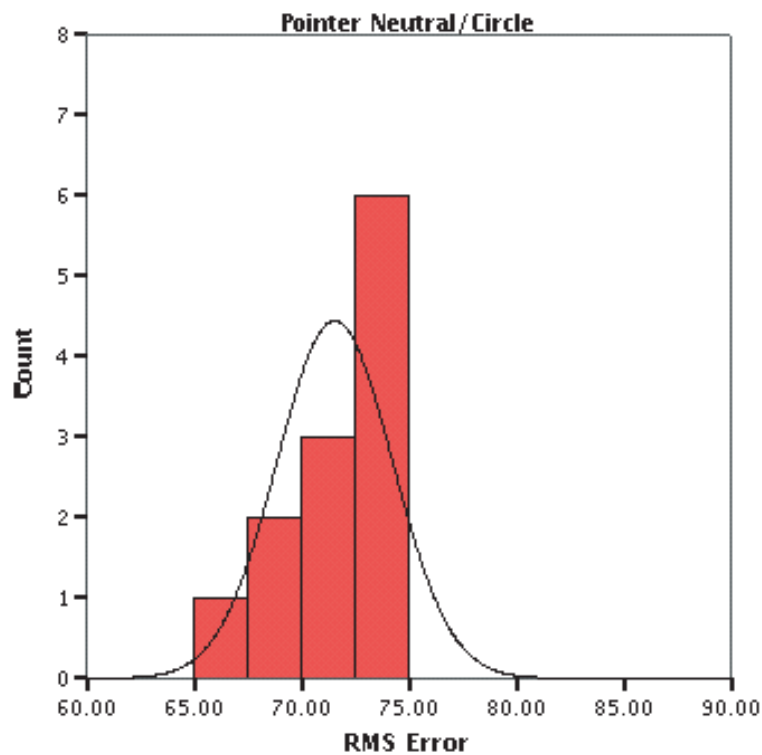


Figure A.10: These histograms present the distribution of RMS errors for each input device and cursor orientation. Each histogram illustrates the distribution of a single cursor orientation and a single input device. The plotted curve indicates the best-fitting normal curve for this particular combination of conditions.

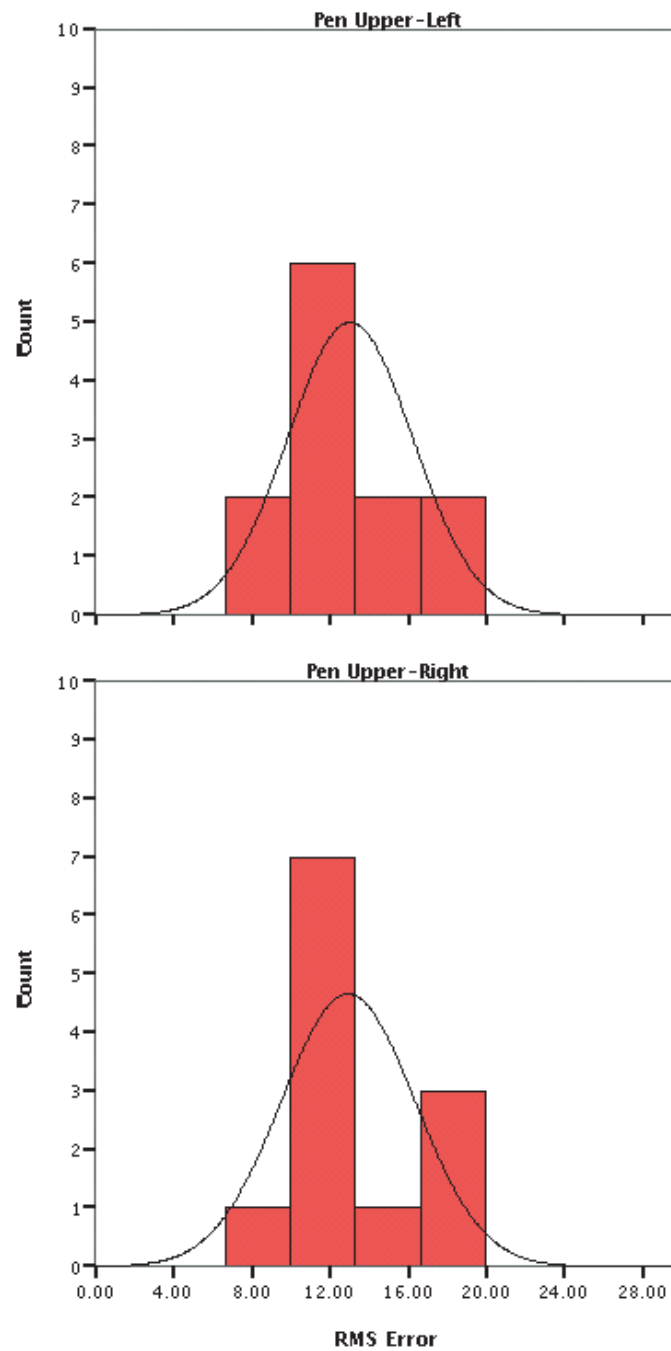


Figure A.11: These histograms present the distribution of RMS errors for each input device and cursor orientation. Each histogram illustrates the distribution of a single cursor orientation and a single input device. The plotted curve indicates the best-fitting normal curve for this particular combination of conditions.

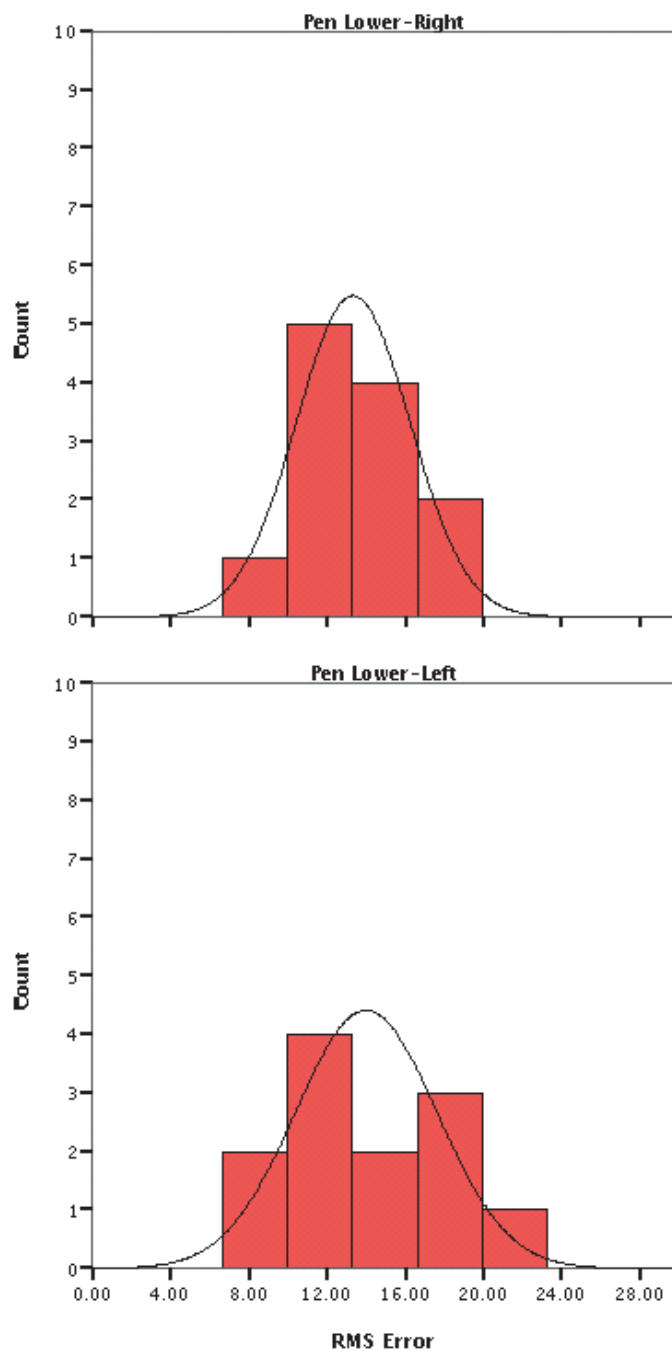


Figure A.12: These histograms present the distribution of RMS errors for each input device and cursor orientation. Each histogram illustrates the distribution of a single cursor orientation and a single input device. The plotted curve indicates the best-fitting normal curve for this particular combination of conditions.

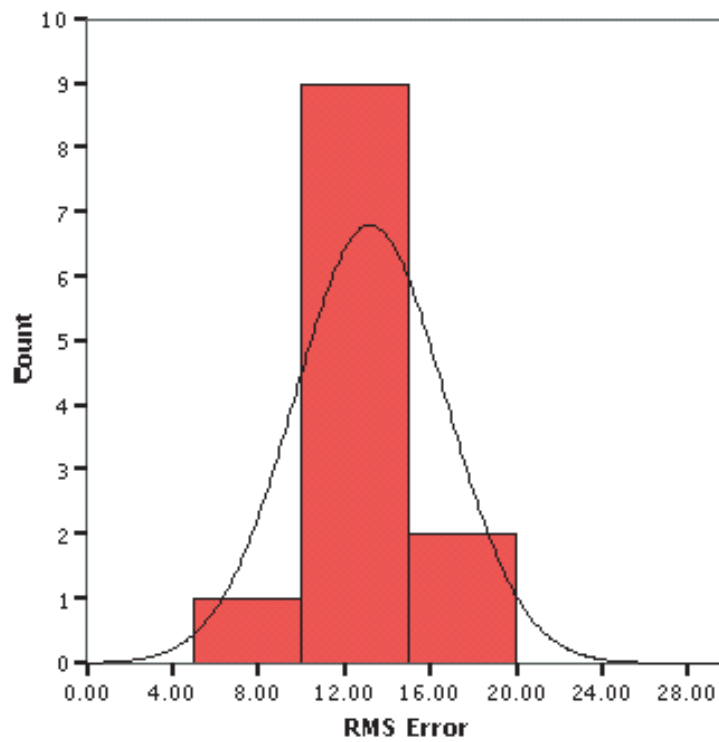


Figure A.13: These histograms present the distribution of RMS errors for each input device and cursor orientation. Each histogram illustrates the distribution of a single cursor orientation and a single input device. The plotted curve indicates the best-fitting normal curve for this particular combination of conditions.

## Appendix B

# Supplemental Data for Experiment

## 2

The following are additional data, which may be helpful in a supplemental analysis of Experiment 2, as presented in Chapter 6.

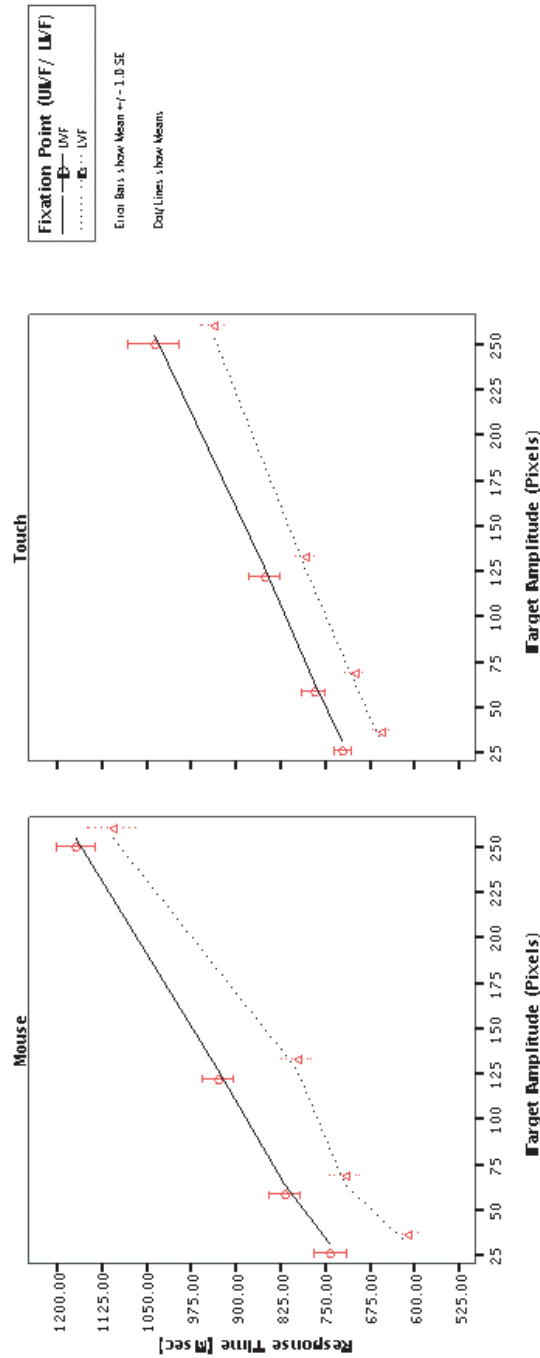


Figure B.1: This line graph plots measured movement time against the experimental factor of target amplitude as an alternative to observing movement time as a function of the index of difficulty (ID). The graph compares UVF and LVF performance across mouse and touchscreen input.

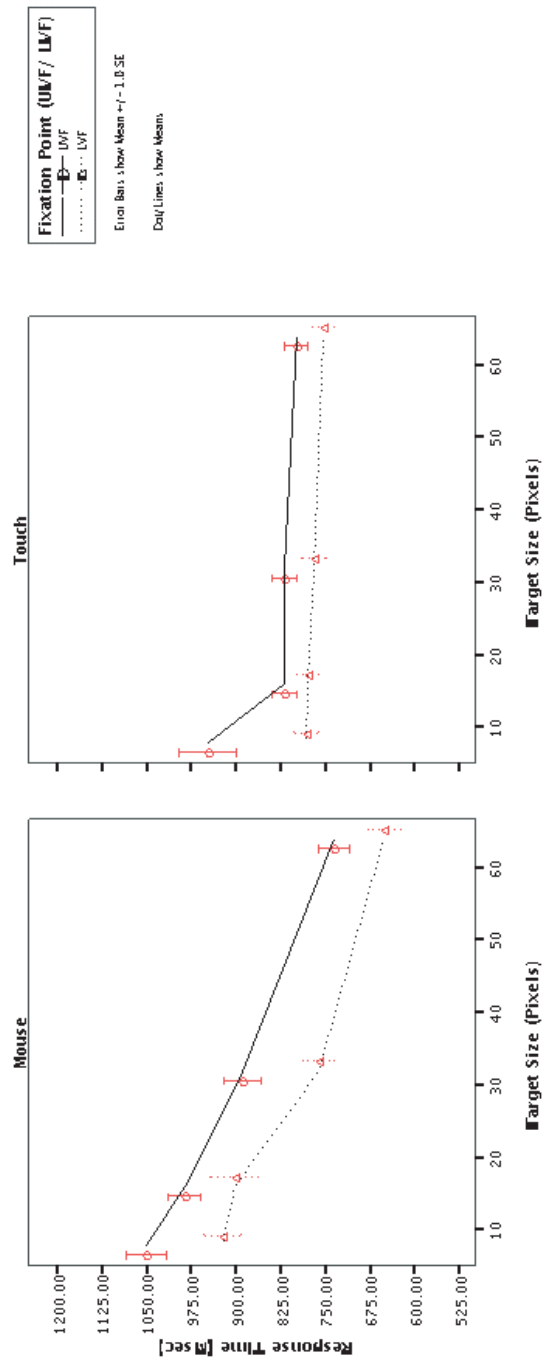


Figure B.2: This line graph plots measured movement time against the experimental factor of target size (width) as an alternative to observing movement time as a function of the index of difficulty (ID). The graph compares UVF and LVF performance across mouse and touchscreen input.

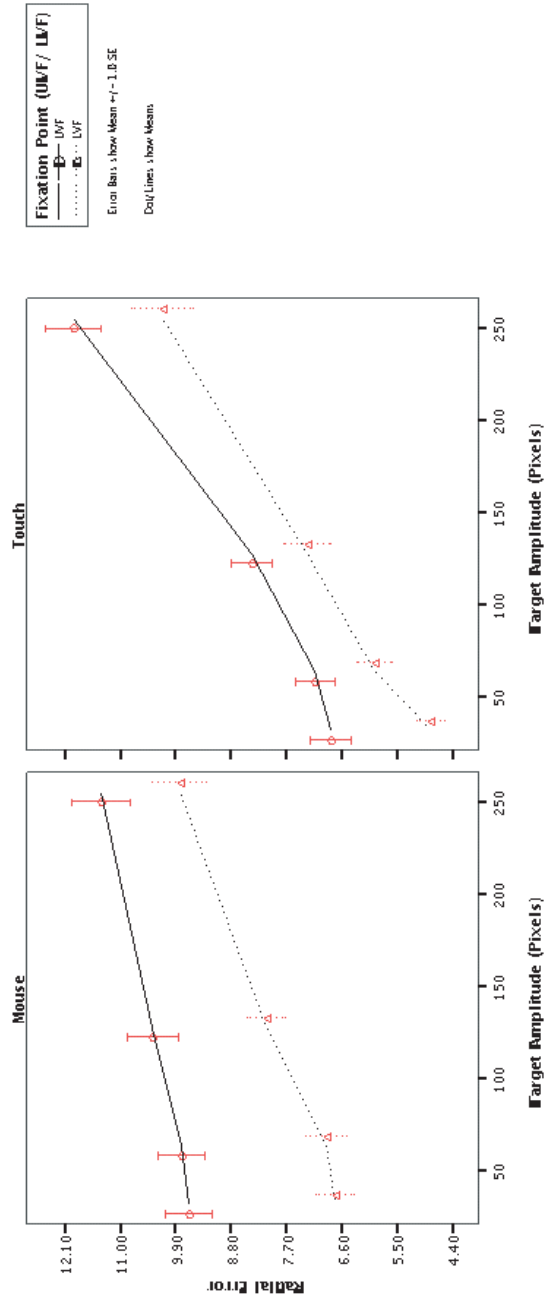


Figure B.3: This line graph plots measured RMS (radial) error against the experimental factor of target amplitude as an alternative to observing pointing precision as a function of the index of difficulty (ID). The graph compares UVF and LVF performance across mouse and touchscreen input.



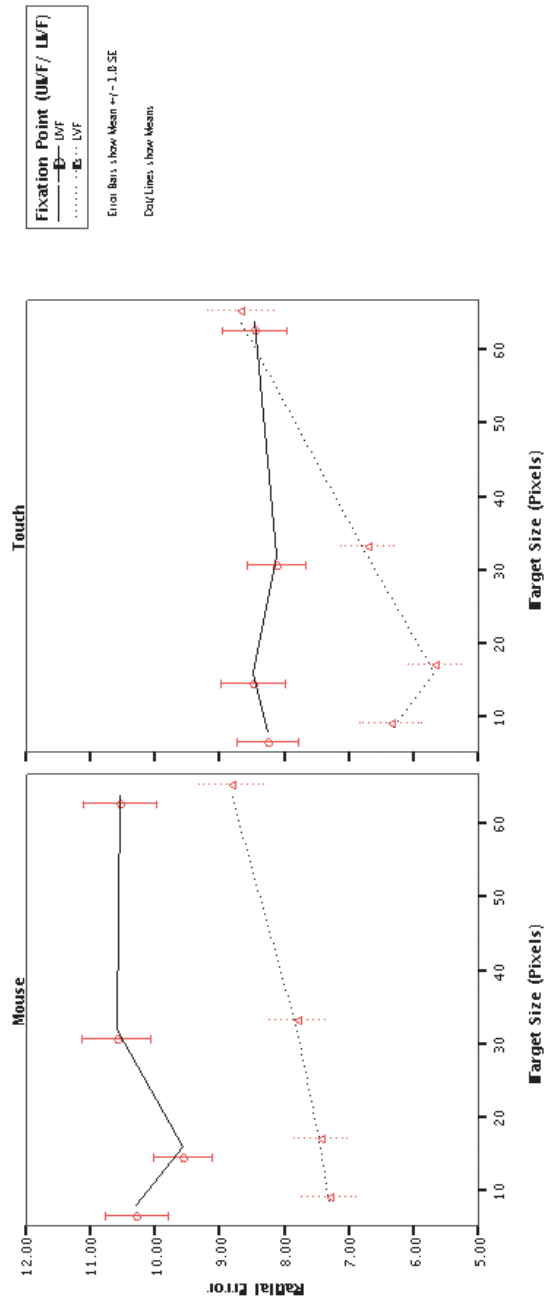


Figure B.4: This line graph plots measured RMS (radial) error against the experimental factor of target size (width) as an alternative to observing pointing precision as a function of the index of difficulty (ID). The graph compares UVF and LVF performance across mouse and touchscreen input.

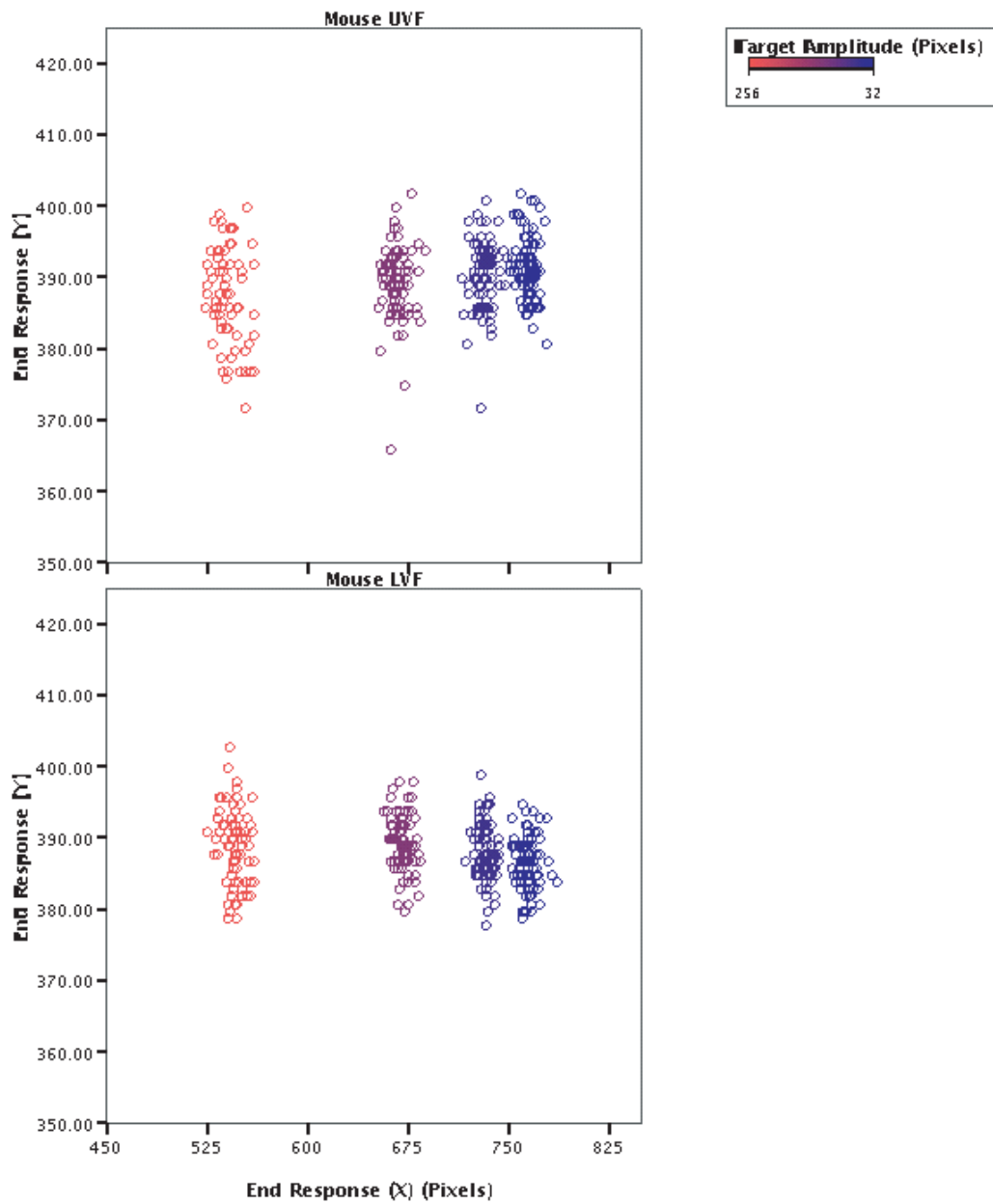


Figure B.5: This scatterplot indicates the locations pointed at by participants as a function of target amplitude. The plot is read from right-to-left, with furthest left indicating the shortest amplitude. The graph compares UVF and LVF performance for mouse input.

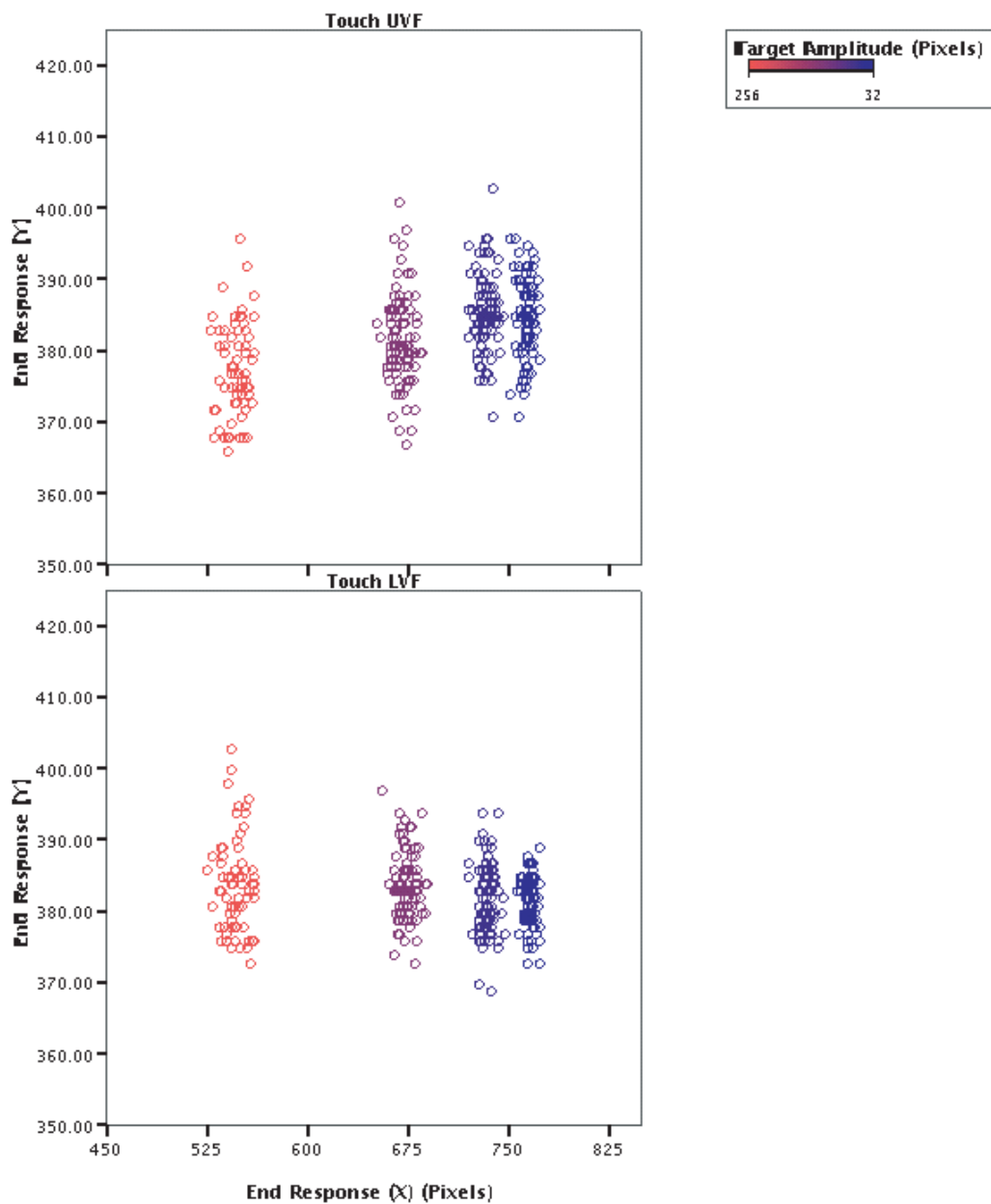


Figure B.6: This scatterplot indicates the locations pointed at by participants as a function of target amplitude. The plot is read from right-to-left, with furthest left indicating the shortest amplitude. The graph compares UVF and LVF performance for touchscreen input.

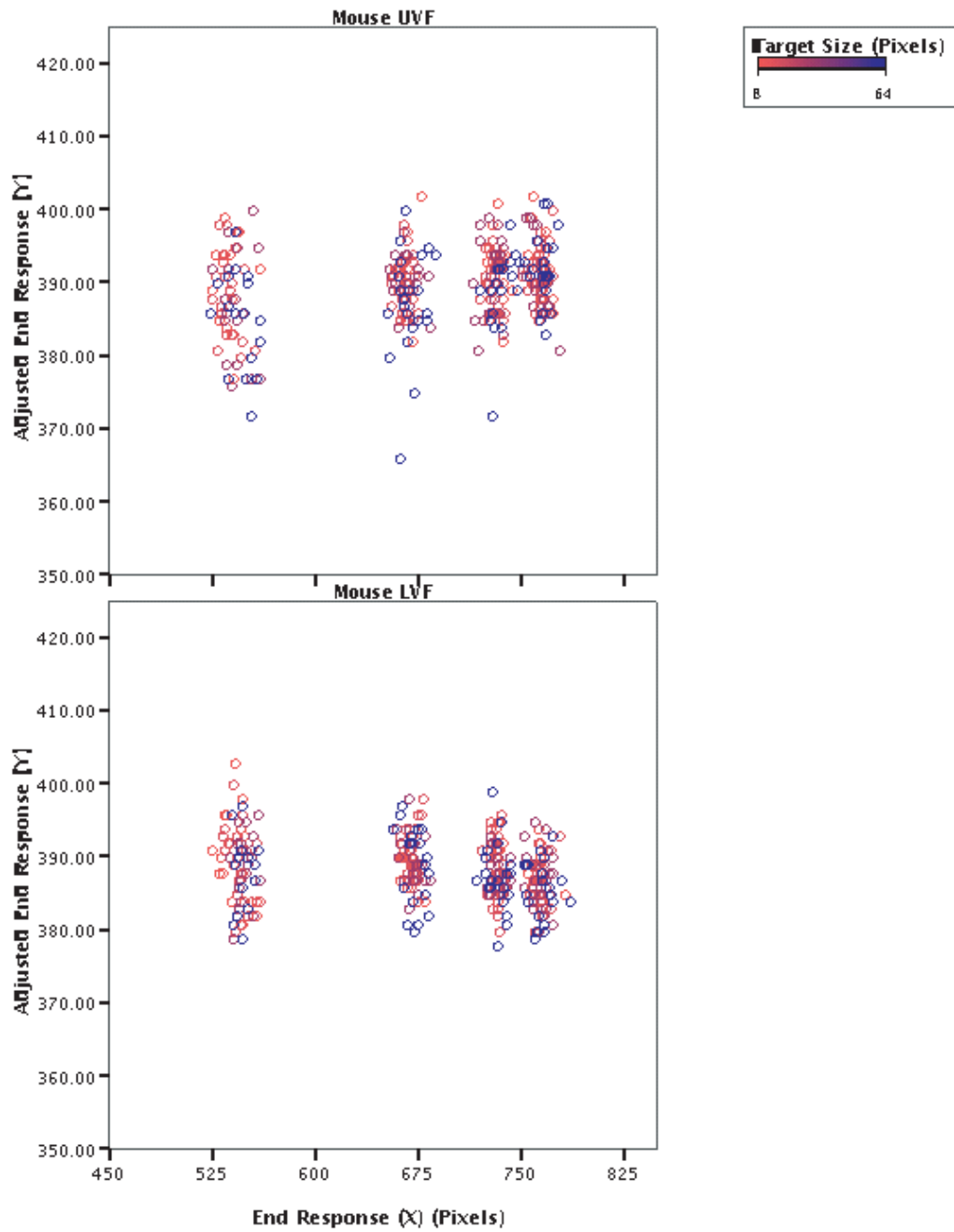


Figure B.7: This scatterplot indicates the locations pointed at by participants as a function of target size (width). The plot is read from right-to-left, with furthest left indicating the shortest amplitude. Target size is differentiated by the spectrum of colours as seen in the figure legend. The graph compares UVF and LVF performance for mouse input.

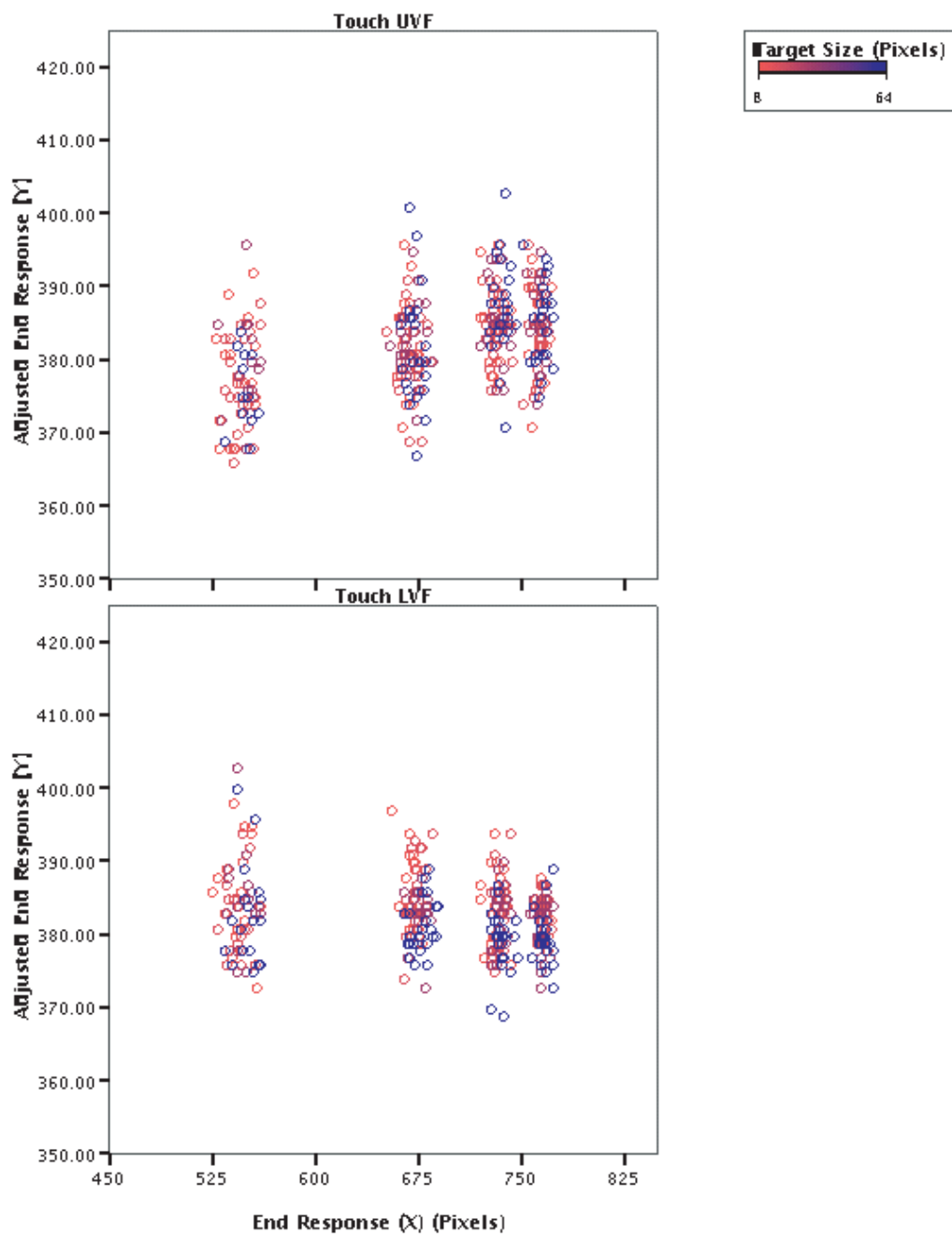


Figure B.8: This scatterplot indicates the locations pointed at by participants as a function of target size (width). The plot is read from right-to-left, with furthest left indicating the shortest amplitude. Target size is differentiated by the spectrum of colours as seen in the figure legend. The graph compares UVF and LVF performance for touch input.

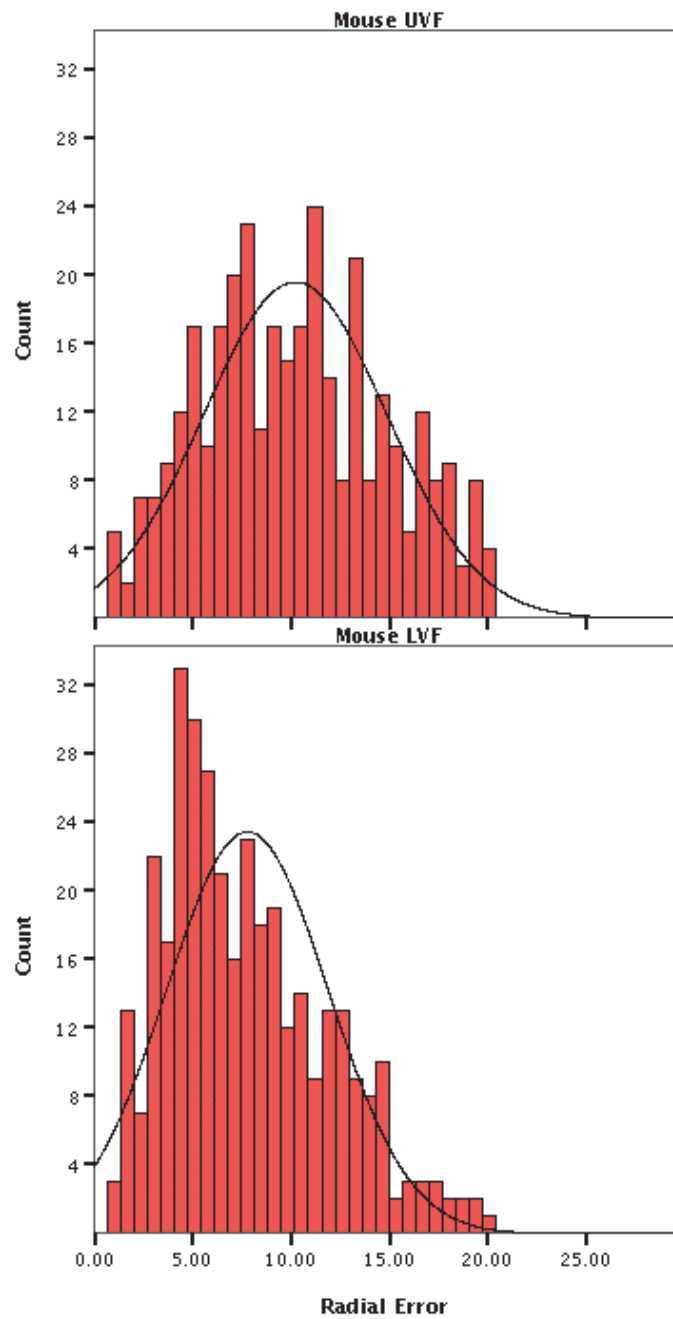


Figure B.9: These histograms present the distribution of RMS errors across the UVF and LVF for mouse pointing. The plotted curve indicates the best-fitting normal curve for this particular combination of conditions.

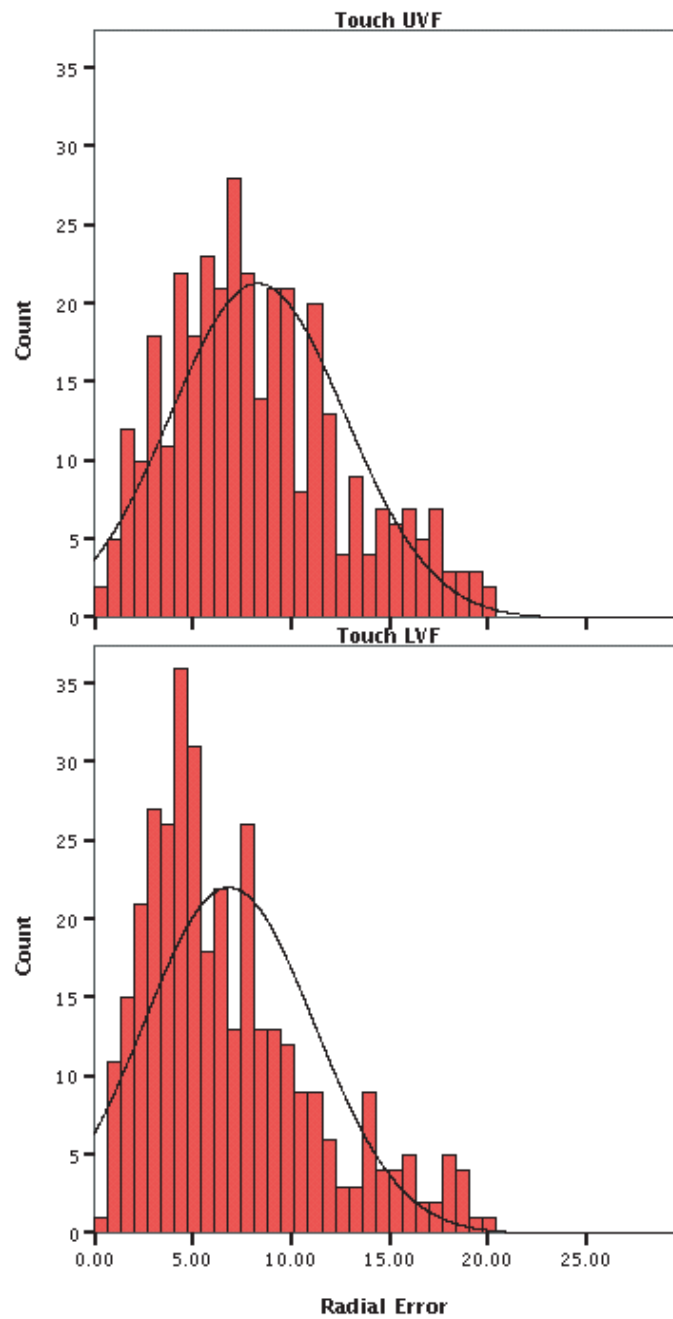


Figure B.10: These histograms present the distribution of RMS errors across the UVF and LVF for touchscreen pointing. The plotted curve indicates the best-fitting normal curve for this particular combination of conditions.





## Appendix C

# Supplemental Data for Experiment

## 3

The following are additional data, which may be helpful in a supplemental analysis of Experiment 3, as presented in Chapter 7.

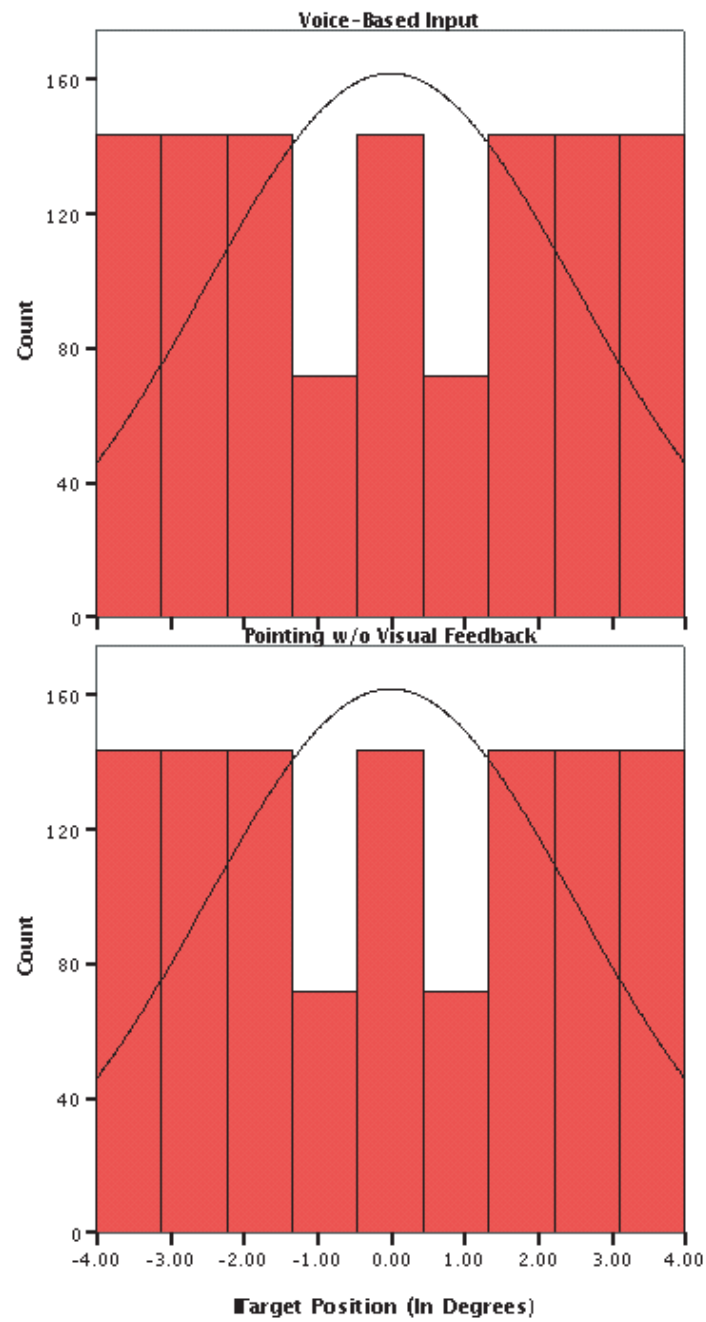


Figure C.1: These histograms show the distribution of randomized target positions as used in Experiment 3. These serve to demonstrate that the trials presented to participants approximated one of a uniform, and not a normal, distribution.

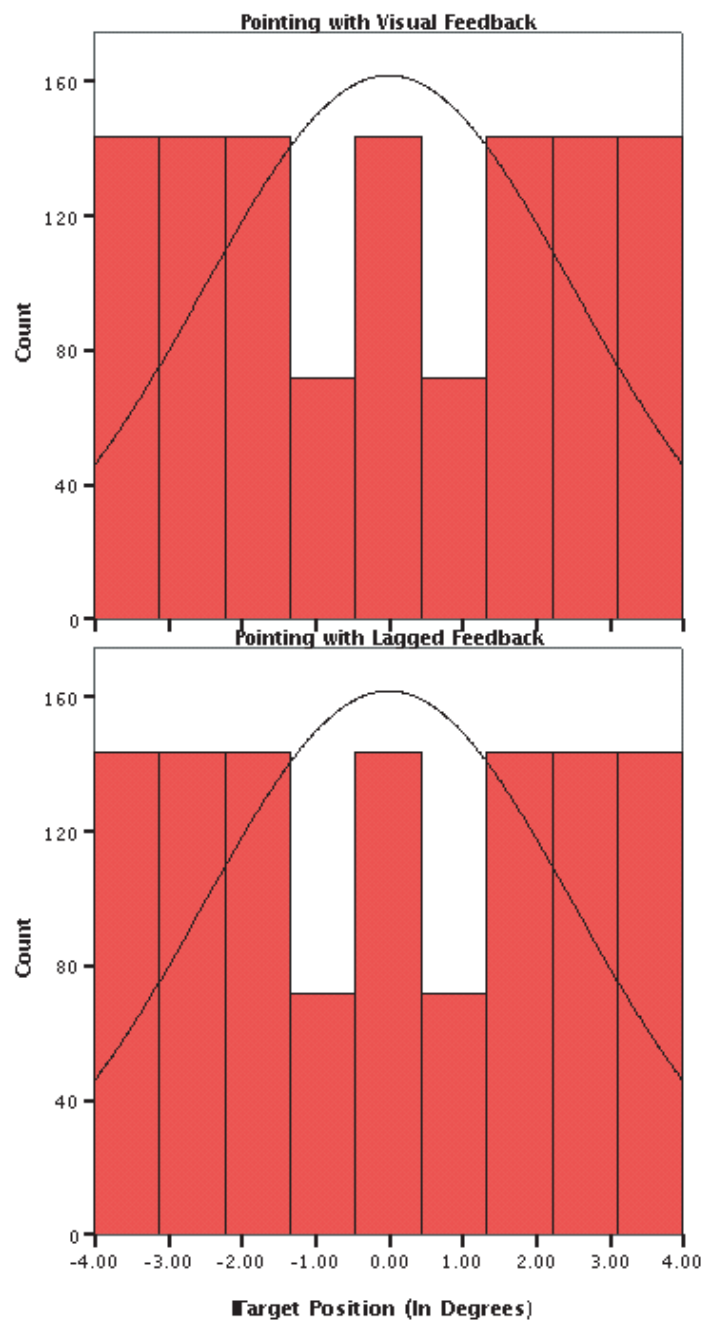


Figure C.2: These histograms show the distribution of randomized target positions as used in Experiment 3. These serve to demonstrate that the trials presented to participants approximated one of a uniform, and not a normal, distribution.

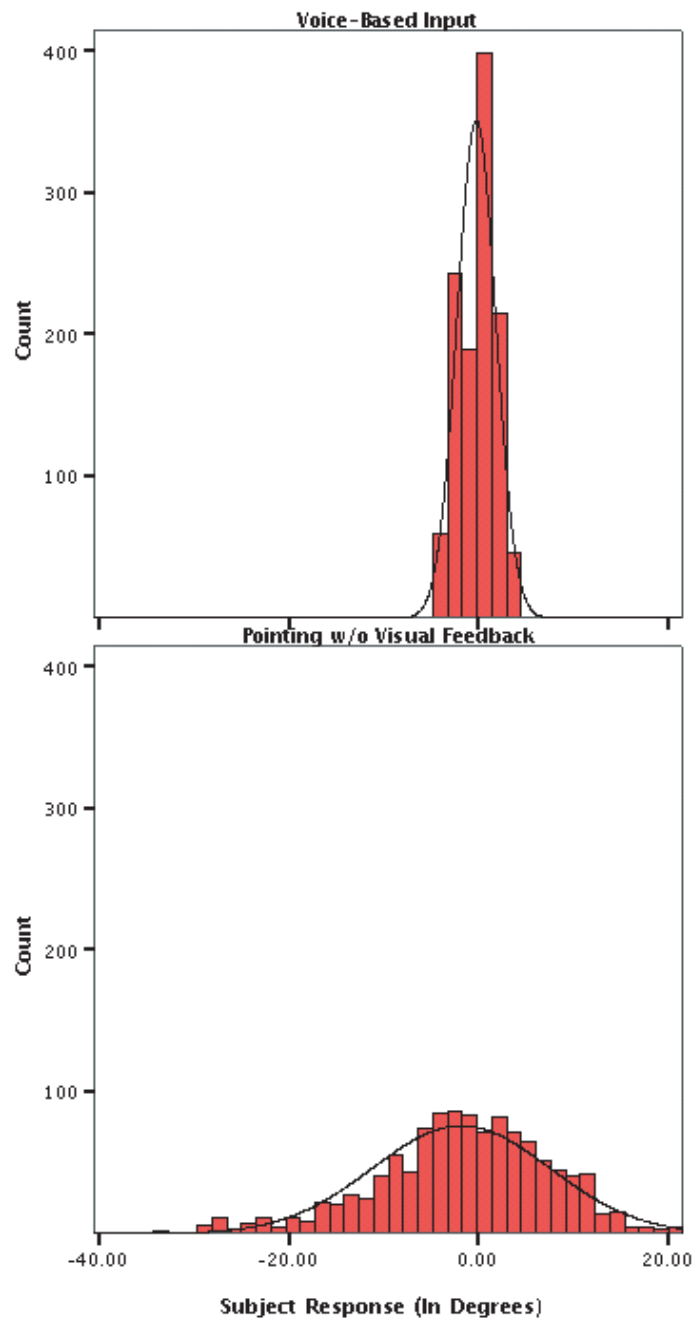


Figure C.3: These histograms show the distribution of pointing errors, with each histogram illustrating the errors for one particular method of input. Participant response was measured along the horizontal axis only.

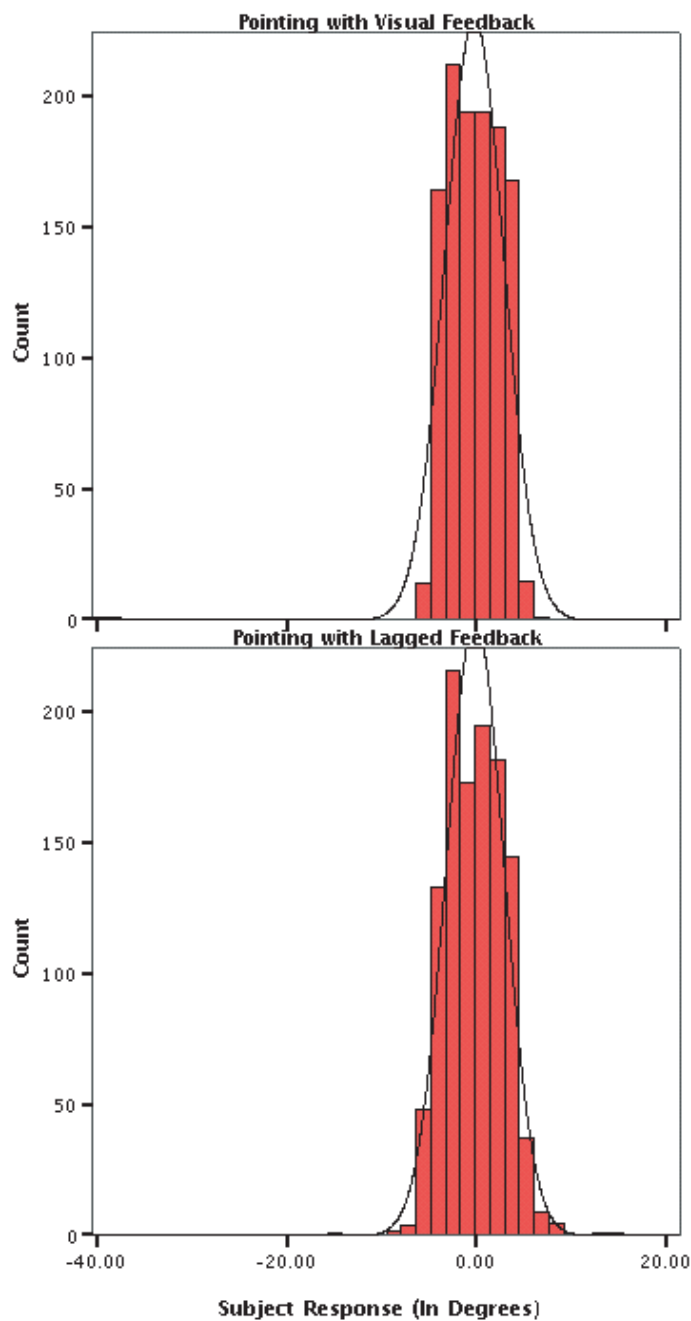


Figure C.4: These histograms show the distribution of pointing errors, with each histogram illustrating the errors for one particular method of input. Participant response was measured along the horizontal axis only.

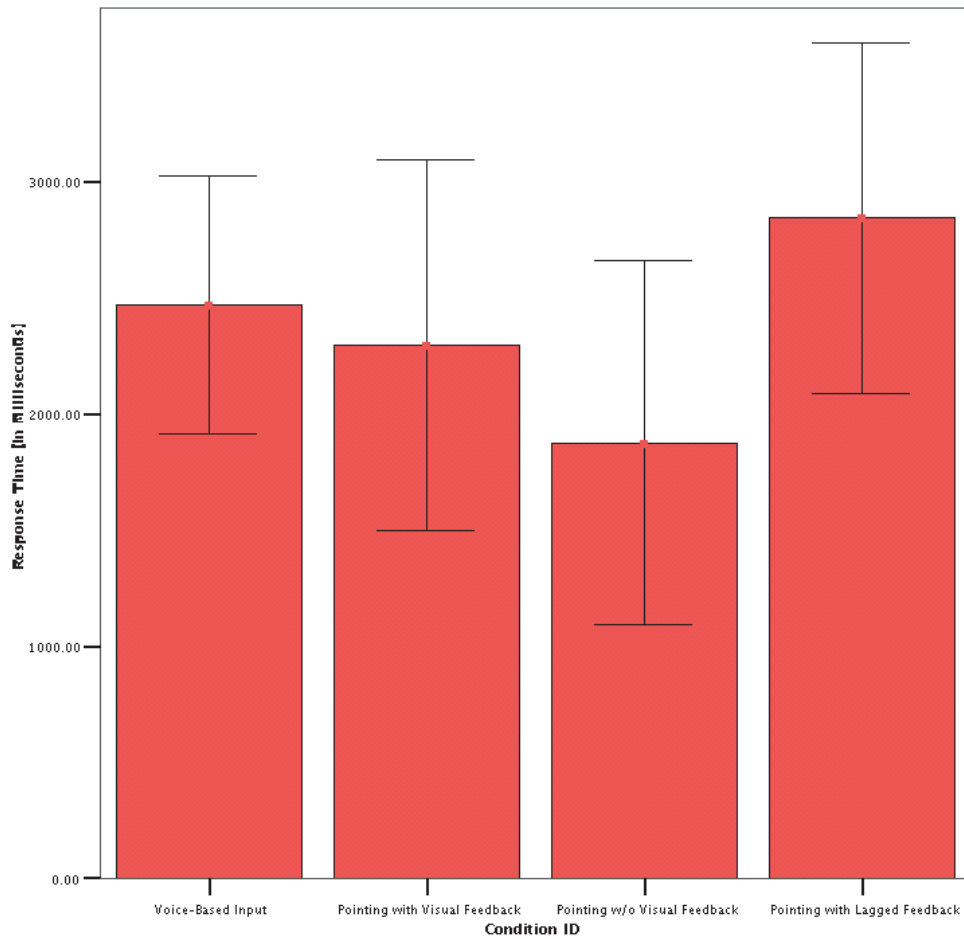


Figure C.5: This bar graph indicates the measured movement times for indicating target positions, broken down by input method.

# Appendix D

## Glossary

This appendix provides a glossary to some of the technical terms, acronyms, and other short-hand forms used in this dissertation. It serves as a useful reference for readers.

**Action** Actor (i.e. user)-generated body movements.

**ANOVA** Analysis of Variance. An inferential statistical technique that is often used in HCI and psychology to evaluate collected quantitative data.

**Coding Hypothesis** An explanation of why stimulus-response (S-R) compatibility effects occur, based on the idea that the compatibility between stimulus and response is related to the complexity of the process by which stimulus information is coded into an appropriate response.

**Computational Theory of Mind** A dominant school of thought in cognitive psychology, which suggests that the human mind can be described as a computational entity and that the behaviour of the mind can be reconciled with its physical implementation in the human brain.

**Cursor** In the context of pointing in HCI, usually refers to a small graphical indicator that tracks movements across a graphical display. These often have directional cues in the form of arrowheads or other similar representations.

**Dorsal Stream** In neuroscience, a portion of the visual pathway dedicated to processing visual information for visually-guided movements and behaviours.

**Fitts's Law** An empirical model of movement time as a function of a log-linear index of difficulty, derived from information theoretic principles. Fitts's Law has been an important tool to understanding how well people can use input devices under various conditions.

**GUI** Graphical User Interface.

**HCI** Human-Computer Interaction.

**Iconic Cursors** Graphical cursors that are used in a visual user interface to indicate the current state of the system.

**Induced Roelofs Effect** A particular visual illusion where a target surrounded by an asymmetric frame leads to a systematic perceptual bias characterized by a perception that the target is further to the left or right than it really is.

**LVF** Lower Visual Field. See Upper and Lower Visual Fields.

**Mental Representation** An abstract structure that encapsulates information about external events in a systematic way so that appropriate decisions may be made.

**Model Human Processor** An information processing model of user performance first introduced by Card, Moran, and Newell (1983). The model attributes user performance as time spent doing a task as a result of the perceptual, cognitive, and motor characteristics of the user.

**Movement Time** In this dissertation, unless otherwise defined, typically refers to the amount of time required for a participant to make a response from the initial onset of a visual stimulus to the time that a final response has been acquired.

**PDA** Personal Digital Assistant. A handheld device that typically uses touchscreen input to allow for a variety of ubiquitous computing tasks.

**Perception** Input to a sensory system (i.e. the human visual system) from the surrounding world.

**Pointing** In HCI, a general term for an entire class of interaction techniques involving a directed motion to various locations on a graphical display.

**Polhemus Fastrak** A six-degree-of-freedom (DOF) spatial input device developed by Polhemus Incorporated that tracks spatial position and orientation using electromagnetic fields. The device is frequently used in virtual reality (VR) settings and large screen, immersive environments.

**Psychophysics** The study of understanding the relationship between objective and subjective events by using a collection of methods meant to separate and characterize these phenomena from each other. This technique is often applied in psychological research.

**Stimulus-Response (S-R) Compatibility** A psychological theory of how visual information might be processed as an integrated construct, which predicts that visual elements like directional cues and their associated responses can influence overall performance (see Chapter 5 for more information).

**Tablet PC** A portable computer, sometimes smaller than a traditional laptop or notebook. It has touchscreen-like capabilities that are enabled via the use of a special pen input device or stylus.

**Two-Visual Systems Hypothesis** A psychological theory of how visual information is processed as a mediated construct. It suggests that visual information is mentally represented and encapsulated in two separate streams of visual processing: a ventral (perceptual) representation and a dorsal (sensorimotor) representation. These lead to distinct response characteristics, and may be driven by the presence or absence of particular visual cues.



**Ubiquitous Computing** A paradigm of user interaction based on the idea that computing devices have become reasonably cheap and affordable so that they can and will be used everywhere and all of the time.

**Upper and Lower Visual Fields** A psychological theory of how visual information might be processed as a separable construct. It suggests that the eye is functionally specialized by where visual activities are most likely to occur. The upper visual field (UVF) is functionally specialized for perceptual tasks while the lower visual field (LVF) is functionally specialized for motor tasks (see Chapter 6 for more information).

**UVF** Upper Visual Field. See Upper and Lower Visual Fields.

**Ventral Stream** In neuroscience, a portion of the visual pathway dedicated to processing visual information for perceptual tasks (i.e. colour, shape, and form).

**WIMP** Windows-Icons-Menus-Pointer. A paradigm for graphical user interface design, common in many desktop settings.