Human Communication Channels in Distributed, Artifact-Centric, Scientific Collaboration

by

Brian D. Corrie
B.Sc., University of Victoria, 1988
M.Sc., University of Victoria, 1990

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Computer Science

Supervisory Committee

Human Communication Channels
in
Distributed, Artifact-Centric, Scientific Collaboration

by

Brian D. Corrie
B.Sc., University of Victoria, 1988
M.Sc., University of Victoria, 1990

Supervisory Committee

Dr. Margaret-Anne Storey, (Department of Computer Science)
Supervisor

Dr. Daniela Damian, (Department of Computer Science)
Departmental Member

Dr. Eric Manning, (Department of Computer Science)
Departmental Member

Dr. Gholamali Shoja, (Department of Computer Science)
Departmental Member

Dr. Rosaline Canessa, (Department of Geography)
Outside Member

Abstract

Supervisory Committee
Dr. Margaret-Anne Storey, (Department of Computer Science)
Supervisor
Dr. Daniela Damian, (Department of Computer Science)
Departmental Member
Dr. Eric Manning, (Department of Computer Science)
Departmental Member
Dr. Gholamali Shoja, (Department of Computer Science)
Departmental Member
Dr. Rosaline Canessa, (Department of Geography)
Outside Member

This dissertation seeks to answer the research questions that arise when digital technologies are used to support distributed, artifact-centric, scientific collaboration. Scientific research is fundamentally collaborative in nature, with researchers often forming collaborations that involve colleagues from other institutions and often other countries. Modern research tools, such as high-resolution scientific instruments and sophisticated computational simulations, are providing scientists with digital data at an unprecedented rate. Thus, digital artifacts are the focus of many of today's scientific collaborations. The understanding of scientific data is difficult because of the complexity of the scientific phenomena that the data represents. Such data is often complex in structure, dynamic in nature (e.g. changes over time), and poorly understood (little *a-priori* knowledge about the phenomena). These issues are exacerbated when such collaborations take place between scientists who are working together at a distance.

This dissertation studies the impact of distance on artifact-centric scientific collaboration. It utilizes a multi-dimensional research approach, considering scientific collaboration at multiple points along the methodological (qualitative/quantitative research methods), cognitive (encoding/decoding), community (many/single research groups), group locality (collocated/distributed), and technological (prototype/production) dimensions. This research results in three primary contributions: 1) a new framework (CoGScience) for the study of distributed, artifact-centric collaboration; 2) new empirical evidence about the human communication channels scientists use to collaborate (utilizing both longitudinal/naturalistic and laboratory studies); and 3) a set of guidelines for the design and creation of more effective distributed, scientific collaboration tools.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AG	AccessGrid	ANOVA	Analysis of Variance
AOI	Area of Interest	API	Application Programmer Interface
CIF	Common Intermediate Format	CREW	Collaboratory for Research on Electronic Work
CSCW	Computer Supported Collaborative Work	DT	Diamond Touch Table
ETNA	Evaluation Taxonomy for Networked Applications	GUI	Graphical User Interface
H261	Video compression protocol	H323	Widely used video conferencing protocol
HCI	Human Computer Interaction	HD	High Definition
HHI	Human to Human Interaction	HSD	Honestly Significant Difference
Hz	Hertz	IRMACS	Interdisciplinary Research in the Mathematical and Computational Sciences
LCD	Liquid Crystal Display	MOC	Mechanics of Collaboration
MPEG	Motion Picture Experts Group	MRT	Media Richness Theory
MST	Media Synchronicity Theory	ms	Millisecond
NGNH	No Gesture No Head	NGYH	No Gesture Yes Head
PARC	Palo Alto Research Centre	RAT	Robust Audio Tool
SFU	Simon Fraser University	SMCR	Source Message Channel Receiver
SMS	Scientific Media Space	SOC	Science of Collaboratories
SYMLOG	System for the Multiple Level Observation of Groups	TIP	Time, Interaction, and Performance
TORSC	Theory of Remote Scientific Collaboration	VAS	Voice Activated Switching
VIC	Video Conferencing Tool	VNC	Virtual Network Computing
WIMP	Windows, Icon, Mouse, Pointer	YGNH	Yes Gesture No Head
YGYH	Yes Gesture Yes Head		

Acknowledgments

After years of research, performing studies, writing collaboration software, and designing, deploying, and operating a wide range of collaboration systems, the list of colleagues, friends, and mentors who contributed to this effort is extensive. To these individuals I owe many thanks. As I will undoubtedly overlook someone important, let me first provide a broad thank you to all who contributed to this research. Without you, this work would have not been possible.

Special thanks go out to my committee. Thanks to Dr. Ali Shoja and Dr. Eric Manning for initially accepting me as their graduate student and for continuing to maintain an interest in my research when my focus changed from network protocols to people. Thanks to Dr. Daniela Damian – from explorations into the impacts of distance on global software development to participation in my gesture study, our discussions were always insightful. Thanks also go to Dr. Rosaline Canessa, my outside committee member, whose perspective as a user of data-centric software tools brought an important perspective to this research. A special thank you to my supervisor, Dr. Peggy Storey. Your mentoring and guidance throughout this process has provided me with the ability to explore this research domain along dimensions that when I started I never would have considered.

Although I primarily worked from afar while performing my research, the Chisel Research Group at the University of Victoria made substantial contributions to my research. Ranging from tough questions at Chisel DesignFests to helping with the design and testing of my gesture research study, help was never far away. In particular, I would like to thank Gargi Boogie for assisting with performing my inter-coder reliability testing and Tricia d'Entremont, Maleh Hernandez, and Peter Rigby for helping me work the bugs out of my gesture study in pre-trial testing. Thanks also to Peter Rigby for giving me a hard time with my statistics.

There are also an extensive number of colleagues external to the University of Victoria who need to be thanked. Although my degree will read the University of Victoria, my research was strongly influenced by the work environment in which I spent my day-to-day life. Much of my early thoughts on frameworks and how to apply them to advanced

collaboration came from work with my colleagues from the National Research Council and Communications Research Centre in Ottawa. I greatly appreciated the many in-depth discussions with Andrew Patrick, Steven Marsh, Janice Singer, Sylvie Noel, Khalil El-Khatib, and Ken Emig. My colleagues within WestGrid, including Jon Borwein, Doug Bowman, Pierre Boulanger, Lyn Bartram, and Kelly Booth, as some of the original architects of the WestGrid collaboration infrastructure, provided me with valuable insight into the world of advanced collaboration environments. Today, these discussions continue with my colleagues from across Canada in the Compute Canada TECC Collaboration Working Group (in particular Scott Wilson, Greg Lukeman, and Leslie Groer). A number of other colleagues at SFU also contributed to this research. In particular, Brian Fisher and Lyn Bartram were always willing to attempt to answer my somewhat naïve statistics questions. Stella Atkins and her research group were extremely helpful in letting me use their Tobii eye tracker as I prepared my gesture study.

A special debt of thanks is owed to my colleagues at the Centre for Interdisciplinary Research in the Mathematical and Computational Sciences (IRMACS). Dr. Peter Borwein, as the visionary who created the IRMACS Centre, recognized the importance of advanced collaboration to the computational science research community a decade ago. It is the implementation of his vision in the IRMACS Centre that created the unique environment that enabled the research performed in this dissertation. Peter, you are truly a visionary in this regard, and it was my great pleasure to work with you in creating the IRMACS Centre. A special thanks to the other two IRMACS “musketeers”, Pam Borghardt and Veselin Jungic, who helped to drive the IRMACS world over the last five years. Never a dull moment! I am also indebted to the rest of the IRMACS team (Glenn Davies, Dominic Lepiane, Doug Johnson, Kelly Gardiner, Andy Gavel, Uwe Glasser, Maryam Elkaswani, Jacob Groundwater, and Reena Rama) who have made the IRMACS Centre such an interesting and exciting place to work! In particular, Andy’s wizardry editing the videos for my gesture study is greatly appreciated.

A special debt of gratitude goes out to Todd Zimmerman. Todd and I originally started working together at the New Media Innovation Centre in 2003, and have been partners in crime throughout the design, implementation, deployment, and operation of both the

WestGrid and IRMACS collaboration infrastructure. Todd, your friendship, knowledge, and collaboration over the past seven years have had a significant impact on this research.

Last, but certainly not least, I have to thank my family and friends for their support. My parents have always been encouraging and supportive of my academic endeavours – even when they come in the form of a mid life crisis that involves starting a PhD at 40 years of age! Rich, if you ever get around to looking at this, you will notice there is nary a teapot to be found. See, I have grown as a PhD student! Cheryl, thanks for providing a bed to sleep in, a barbecue to cook burgers on, and a fridge to keep the beer cold during my many trips to Victoria.

Of course my main supporters in this effort have been my wife Sherri and my daughter Lorissa. Without your unending support, encouragement, and prodding, this would not have been possible. Thank you for your patience and love. And yes, Lorissa, I think my PhD is done...

Dedication

For Sherri and Lorissa

I owe you one!

1 Introduction

This dissertation seeks to answer the Computer Supported Collaborative Work (CSCW) research questions that arise when digital technologies are used to support distributed, artifact-centric, scientific collaboration.

How people communicate has been studied since antiquity, with some of the early known published works going back to Aristotle's and Cicero's treatises on the art of rhetoric and oratory. Since that time, contributions to research on human communication and collaboration have come from a wide range of scientific disciplines, including sociology, psychology, linguistics, and communication. In particular, the study of how groups work together has been the target of intensive research for over a century [PS99], utilizing both theoretical and empirical methods to create a range of theories, models, and frameworks on how humans communicate and how groups work together. Over the past thirty years, the widespread use of computers, the Internet, email, and video conferencing have had a dramatic impact on how people work together. Globally distributed work groups are rapidly becoming the norm, rather than the exception.

This trend towards globalization is clearly present in the academic research community. Scientific research is fundamentally collaborative in nature, and many of today's scientific problems require domain expertise in a wide range of disciplines. In order to explore such problems, researchers form collaborations that involve colleagues from other institutions, often located around the world. Modern research tools, such as high-resolution scientific instruments and sophisticated computational simulations, are providing scientists with data at an unprecedented rate. Thus, the focus of many of today's scientific collaborations is on digital data. The understanding of such data is particularly difficult because of the complexity of the scientific phenomena that the data represents. The data is often complex in structure, dynamic in nature (e.g. changes over time), and poorly understood (little *a-priori* knowledge about the phenomena is available). These issues are exacerbated when such collaborations take place between scientists who are working together at a distance. *The focus of the research presented in this dissertation is on the impact that distance has on distributed, data-centric, scientific collaboration.*

1.1 Motivation

Computing is about insight, not numbers.

Richard Hamming, 1962

Over the past fifty years, scientific research has been profoundly impacted by the rapid change in technology. Computational science is the domain of scientific research in which the computer is one of, if not the key, scientific research tool. Computational science complements, supports, and extends the traditional experimental and theoretical approaches to scientific investigation. The dramatic increase in the amount of data that is available to scientific researchers, using high-resolution instruments and/or increasingly complex computational simulations, is transforming the way scientists perform research.

Richard Hamming's insightful statement that *"Computing is about insight, not numbers"* [Ham62] anticipated the problems to which this data deluge would lead. Although computational simulation, data reduction techniques, data mining, and knowledge extraction are all important tools to today's computational scientist, ultimately insight comes from the scientist's interpretation of the data. Thus, the collaborative exploration of *digital artifacts*[1] that represent complex scientific phenomena is becoming an increasingly important tool to the scientific research community.

This problem domain is a complex one, requiring knowledge and understanding in the areas of sociology and group work, cognitive psychology and perception, communication theory, gestural communication, human computer interaction, digital media, advanced networking, and CSCW. Although current literature provides a number of theories, models, and frameworks that attempt to capture this complexity, a sufficiently comprehensive and cohesive framework that brings these fields together has been elusive.

1.1.1 Personal Motivation

This research is driven by two key personal experiences, one inspirational and one opportunistic. The *inspiration* for much of this research comes from experiences gained

[1] We define a digital artifact as any collection of digital data that is displayed on a computer screen in a manner that allows it to be identified as an individual entity. That is, an artifact is a visual representation of an abstract or concrete concept (displayed on a computer screen) that can be identified, pointed at, or acted on.

through the creation and use of a prototype advanced collaboration environment. The CoTable system (described in Chapter 5) utilizes a touch-sensitive tabletop interaction technology that enables the exploration of rich, multi-modal, artifact-centric collaboration. Experiences in designing, building, and experimenting with this system made two things immediately clear. First, in applying existing theories, models, and frameworks to the design and implementation of CoTable, it became clear that no one framework covered the rich interactions that we needed to capture. In order to capture these subtleties, a new framework was required. Second, the rich, multi-modal interactions that were enabled in the CoTable system resulted in complex, and seemingly counterintuitive interactions occurring between users. In particular, the way gestural interaction was utilized by the users of the CoTable system raised many intriguing questions about gesture and its use in artifact-centric, distributed collaboration. The questions raised in developing and experimenting with the CoTable system eventually led to a topic change in this research. What was initially a research focus on the networking issues of advanced collaboration (network protocols, video compression) gradually became a focus on CSCW, Human-Computer Interaction (HCI), and social psychology. Experiences with CoTable inspired much of the research carried out in this dissertation, and the influence of the questions raised through the use of the CoTable system can be seen throughout the research objectives listed in Section 1.2.

The *opportunity* that enabled much of this research is in fact not directly research related, but instead related to sustenance. In November 2004, the New Media Innovation Centre, my employer at the time, closed its doors. In December 2004, I began a new position, continuing some of the work that I was carrying out at the New Media Innovation Centre. This new position was as the Collaboration and Visualization Coordinator for two large research organizations[2]. WestGrid is a large, multi-university computational science consortium that spans all of the major universities in Western Canada. The Centre for Interdisciplinary Research in the Mathematical and Computational Sciences (IRMACS) is a large interdisciplinary research facility at Simon Fraser University. My role for both of these organizations was to design, develop, deploy, and operate an advanced collaboration and visualization infrastructure for scientific

[2] My day job, so to speak.

research. This included the coordination of the design, implementation, and operation of the facilities that support distributed scientific collaboration across the fourteen WestGrid institutions. Could there be a better environment in which to perform research into understanding the collaboration needs of scientific researchers?

This convergence of inspiration and opportunity are two of the key motivating factors that have driven this research. The detailed objectives, methodology, scope, and contributions of the research are elaborated on below.

1.2 Research Objectives

Bringing researchers together to explore the complex phenomena common in today's computational science, and ultimately to accelerate scientific insight, is the lofty goal of this research. We take significant steps toward reaching this goal by pursuing the following objectives:

Objective 1: Develop a broad understanding of how scientific researchers collaborate.

Objective 2: Develop a deep understanding of how scientific researchers interact with digital artifacts when they collaborate.

Objective 3: Evaluate advanced collaboration modalities and technologies for scientific collaboration.

Objective 4: Develop a set of design guidelines for the development of effective collaboration tools for scientific researchers.

In particular, these research objectives naturally lead to the following research questions:

Objective 1: Develop a broad understanding of how scientific researchers collaborate.

1. *How do collaboration patterns change in the presence of technology?*

Objective 2: Develop a deep understanding of how scientific researchers interact with digital artifacts when they collaborate.

1. *What role do digital artifacts play in scientific collaboration?*
2. *What information is lost when such collaboration takes place at a distance?*
3. *What communication channels are used to encode information during artifact-centric collaboration?*

4. *What communication channels are used to decode information during artifact-centric collaboration?*

Objective 3: Evaluate advanced collaboration modalities and technologies for scientific collaboration.

1. *How do researchers use advanced collaboration technologies?*
2. *How well do those technologies work*?

Objective 4: Develop a set of design guidelines for the development of effective collaboration tools for scientific researchers.

1. *What human communication channels need to be supported for artifact-centric collaboration?*

1.3 Approach and Methodology

This dissertation contributes new knowledge and new research tools in the area of distributed, artifact-centric, scientific collaboration. In particular, this research focuses on where the social and cognitive aspects of artifact-centric collaboration intersect with the human-computer interaction and computer supported collaborative work domains of computer science. This dissertation accomplishes this using several different, but complimentary research approaches:

- It utilizes both quantitative (laboratory experiments) and qualitative (ethnographic/naturalistic) research methods to perform the above analysis. Multiple methods are used within studies as well as across the four studies carried out as part of this research.
- It analyzes the use of advanced collaboration tools at the macro-level (use by researchers at a large research centre over a five year period) and the micro-level (use of advanced artifact-centric collaboration tools by a single research group over a five-month period).
- It analyzes the use of both collocated (collaborators in the same room) and distributed (collaborators at two or more distributed locations) collaboration.
- It analyzes both the encoding (how information is sent) and decoding (how information is received) processes researchers use to communicate about complex scientific topics.

- It analyzes the use of state-of-the-art technical infrastructure in both research prototype (experimenting with new HCI and CSCW technologies) and production (observing active researchers using sophisticated CSCW tools) environments.

This dissertation uses the above analyses to identify several key problem areas in artifact-centric collaboration. In particular, given the domain of artifact-centric collaboration, it provides a detailed analysis of the human communication channels used in both collocated and distributed scientific collaboration and the impacts that distance has on the effective communication of interactions with complex digital artifacts.

1.4 Scope

Collaboration, like the term "group work", can be used to describe almost any human interaction that entails trying to accomplish a task. Thus it is critical to define precisely what is meant by *distributed, artifact-centric, scientific collaboration*.

Collaboration is typically categorized on two dimensions, time and place [Bae93, p. 3]. The time dimension captures whether the collaboration takes place at the same time (synchronous collaboration) or over an extended period of time (asynchronous collaboration). The place dimension captures the geographic distribution of the participants. This distribution can be complex, ranging from all participants being collocated at the same physical location, through two or more groups of varying sizes being distributed geographically, to many individual participants all of whom are geographically distributed. This dissertation focuses on *synchronous* collaboration, but considers a range of distribution scenarios, studying both *collocated* and *distributed* collaboration.

This research is also focussed on *scientific collaboration*. This focus is driven by two key factors. First, distributed research teams are rapidly becoming the norm, yet collaboration tools that meet the specific needs of collaborative science are rare. Second, the CSCW community has not explored scientific collaboration in great detail, and there is an opportunity to make a significant impact in this area.

Lastly, this research focuses on how digital data, or *artifacts*, are used by scientists as part of their collaboration. We come back to Hamming's statement, *"Science is about insight, not numbers"* [Ham62]. Driven by the deluge of data that is being produced by advanced scientific instruments and computational simulations, the creation of effective

artifact-centric collaboration tools has the potential to accelerate the researcher's path to such insight. In particular, the use of gesture and how it is used to interact with complex digital artifacts is a key focus of this research.

1.5 Contributions

This research contributes to the group work and scientific collaboratory research communities through the creation of the CoGScience Framework, a new framework for studying distributed, artifact-centric, scientific collaboration. The development of the framework was driven by the realization that current theories, models, and frameworks did not adequately describe distributed, artifact-centric, scientific collaboration at the level of detail required for this research. As early studies were carried out, current frameworks were extended to incorporate theoretical concepts that were relevant to this research but did not exist in any single framework. These extensions were based on relevant theory from cognitive psychology, communication theory, sociology, and group work. Using it as a lens with which to analyze data-centric collaboration, the CoGScience Framework provides a new method for comparing past research, analyzing existing collaboration tools, designing new research studies, and designing new collaboration environments.

This research contributes new empirical evidence to the CSCW and social psychology communities. The empirical results presented in this dissertation add new knowledge about how scientific researchers interact with digital artifacts and how that interaction is impacted when researchers are at a distance. In particular, our longitudinal (five month) study of a working research group in a naturalistic, yet high technology, collaboration environment provides us with a unique perspective on how scientists collaborate. The results presented here also add new knowledge to the social psychology community on how human communication channels are used to both encode and decode information when researchers are interacting with digital artifacts. In particular, our approach on using eye tracking to analyze the decoding process of human communication is, to our knowledge, unique.

Finally, this research synthesizes the results from the studies presented in this dissertation into a set of design guidelines for the creation of effective, artifact-centric

collaboration tools for the scientific research community. These design guidelines are targeted at both tool developers and infrastructure designers and operators.

1.6 Evaluation

The CoGScience Framework is validated by analyzing its efficacy at capturing the key details of the research studies presented throughout this dissertation. It is used to perform top-down analyses of a number of different collaboration tasks, a bottom-up analysis of several advance collaboration systems, a comprehensive analysis of a distributed, tabletop collaboration prototype, and the design and analysis of an experimental laboratory study.

The empirical results are evaluated in terms of their effectiveness in meeting the research goals and objectives. That is, do our empirical results provide new evidence that helps to answer the research questions and objectives? Do our results support or refute existing theory? Do our results help to provide practical guidelines for creating effective artifact-centric collaboration tools for the scientific research community? This evaluation is done at both the theoretical and practical levels.

1.7 Organization of this Dissertation

This dissertation is organized in four parts. Part 1 (Chapter 2) explores the wide range of research domains that influence data-centric, scientific collaboration. This includes relevant research in the computational sciences, the social sciences, and computer science. Part 2 (Chapter 3 and Chapter 4) considers the methodological aspects of this research. Chapter 3 discusses the research methodology used in this dissertation. Chapter 4 describes the CoGScience Framework, a methodological tool developed as a key component of this research. Part 3 (Chapter 5 through Chapter 11) presents the studies carried out as part of this dissertation. Chapter 5 describes the creation of the CoTable collaboration environment and our experiences with its use. Chapter 6 presents a case study that analyzes how the installation and support of state-of-the-art distance collaboration tools have changed the collaboration pattern of researchers at a large research centre over a five year time period. Chapter 7 presents a naturalistic, ethnographic study that analyzes the usage of advanced distance collaboration tools by a single research group over a five month period. Chapter 8 through Chapter 11 present a

laboratory study that analyzes the decoding process (how information is processed) used by researchers during scientific presentations. Part 4 (Chapter 12 and Chapter 13) summarizes the results of this research. Chapter 12 aggregates the knowledge gained across the research presented in the other chapters, coalescing the information into a set of design guidelines for the creation of effective collaboration tools for the computational sciences. Chapter 13 provides an overview of how the research presented in this dissertation meets the research objectives listed above, describes the contributions that result from this research, discusses areas for future research, and draws some final conclusions.

2 Related Research

This chapter presents an overview of the foundational research areas that are necessary to understanding the domain of distributed, artifact-centric, scientific collaboration. First, we discuss current research into collaboration in the sciences, considering the domain of computational science, scientific collaboratories, and data-centric science. We then explore the science of collaboration, considering a broad range of related research areas, including communication, social psychology, language use, gesture, and cognitive psychology. This is followed by a discussion of the related research in the domain of Computer Supported Collaborative Work. Lastly, we consider how all of these domains are inter-related by exploring theories, models, and frameworks that tie this research together.

2.1 Collaboration in Science

2.1.1 Scientific Collaboratories

Over the last twenty years, large scale distributed research groups, or collaboratories (as originally coined in 1989 by Wulf [Wul89]), have become common in many areas of science [BZO+07]. The US National Research Council's report on collaboratories [NRC1993] defines a collaboratory at the abstract level, using Wulf's terminology, as a *"...center without walls in which the nation's researchers can perform research without regard to geographical location, interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information from digital libraries."*

Collaboratories and the related scientific research infrastructure have been explored in some detail in the recent research literature. The Science of Collaboratories (SOC) project, based at the University of Michigan, has conducted a broad review of a wide range of collaboratory projects [OZB08]. One of the important research outcomes from this work is the creation of a taxonomy of collaboratory types [BZO+07, BZO+08]. They classify collaboratories based on the focal point of the collaboration. These focal points are:

- Shared Instrument: A collaboratory that provides remote access to expensive scientific instruments such as a telescope or particle accelerator.
- Community Data Systems: A collaboratory that is formed around a common data archive.
- Open Community: A collaboratory that aggregates the expertise of many people towards solving a specific problem.
- Virtual Community of Practice: A group of people who share a research area and communicate about it online.
- Virtual Learning Community: A community whose goal is to increase the knowledge of participants (but not necessarily perform research).
- Distributed Research Centre: A distributed group of people, equipment, and resources that work together on a set of joint projects.
- Community Infrastructure Project: A set of infrastructure (software tools, protocols, instruments, computers) that facilitates science.

Finholt has also recently explored a wide range of scientific collaboratories, attempting to identify factors that can help to predict the success and failure of such organizations [Fin03]. He points out that the social and behavioural aspects of collaboratories may be as important, if not more so, than the traditional collaboratory focus on remote access to data and/or observation and operation of scientific instruments. Finholt states "*… the critical element of collaboratories – for scientists – might be the opportunity they allow for encounters, discussions, and sharing of ideas*."

Another important dimension in the exploration of scientific collaboratories is the relative lack of rigorous analysis of collaboratory success. Sonnewald *et al.* point out that the evaluation of scientific collaboratories has lagged behind the development of the infrastructure [SWM03]. For example, in *Scientific Collaboratories on the Internet* [OZB08], the most comprehensive book to date on scientific collaboratories, there is only one chapter that presents a quantitative evaluation of scientific collaboration [SWM08]. In this paper, which is an extension of their 2003 study [SWM03], the authors raise a number of questions that are highly relevant to this research, including:

- How does the scientific process change in the context of a collaboratory?
- Will scientists adopt collaboratory software?

- How do organizational cultures impact adoption of collaboratory systems?
- Are there system features and performance characteristics that are common to successful collaboratory systems?

In fact, several of these questions are reflected in the research questions posed in this dissertation.

Sonnewald *et al.* attempt to answer some of these questions, performing a laboratory study on the nanoManipulator collaboratory [SWM03, SWM08]. The study shows that there is no statistical difference in the performance of face-to-face and distributed scientific teams on the collaboration task explored. In fact, participants in the study stated that there were benefits and drawbacks to both collocated and distributed collaboration. Such results are important, as to date there has been little quantitative analysis of distributed scientific collaboration.

Although not focused on distributed collaboration, two other relevant research projects that explore collocated scientific collaboration are worthy of note. Huang *et al.* performed a post-hoc analysis (through interviews with scientific staff who used the system) of the use of the MERBoard system, a large screen collocated collaboration environment designed for the Mars Exploration Rover (MER) mission [HMT06]. Wigdor *et al.* performed a participatory design and evaluation of a collocated large screen tabletop and wall display system called WeSpace [WJF+09]. Although these analyses are qualitative in nature, they are rare in that they explore the use of advanced technologies in a naturalistic scientific environment. Because both of these systems are collocated collaboration systems, we consider them in more detail in Section 2.3.1.4.

There is clearly much research that remains to be performed this area. The research presented in this dissertation adds both new qualitative and quantitative results to this domain.

2.1.2 Data-Centric Science

Of particular relevance to this research is the work of Birnholtz *et al.* on the role data plays in scientific collaborations [BB03]. The authors argue that in order to develop collaboration tools that support data sharing, a better understanding of how researchers use data is required. Their research suggests that data plays two main roles in scientific communities: the widely recognized role of providing evidence to support scientific

inquiry and the less obvious role of making a social contribution to the establishment and maintenance of communities of practice. In particular, their analysis describes how data use defines boundaries between scientific approaches (experimental and theoretical), how access to data acts as a gateway to a community of practice (if you can access the data, you are part of the community), how access to or ownership of data brings status to researchers or research groups, and how access to data can enable more extensive participation in research communities. One of the fundamental suggestions the authors make is that it is necessary for a collaboratory to support both the scientific and social roles that data plays in a community of practice.

As collaboratories continue to emerge, collaboration tools that support distributed, data-centric research will continue to increase in importance. It is clear that we need a much better understanding of the role data plays in such collaborations.

2.1.3 Scientific Visualization

Scientific visualization is the process of making images from scientific data for the purpose of increasing the understanding that we have about that data. Our visual system provides the highest bandwidth channel from the computer display to our brains [War04, p. 2]. Studies have shown that the human visual and cognitive systems are adept at detecting patterns in data, helping individuals make inferences about data, and helping them form hypotheses about that data [CMS99, War04]. Given the rapid growth of the size and complexity of the data sets from today's computational simulations and high resolution scientific instruments, the processing and understanding of complex scientific data is rapidly increasing in importance [JMM+06]. In their 2006 report on the Research Challenges in Visualization to the US National Science Foundation and National Institute of Health, Johnson et al. state that *"Visualization is indispensable to the solution of complex problems in every sector, from traditional medical, science and engineering domains to such key areas as financial markets, national security, and public health. Advances in visualization enable researchers to analyze and understand unprecedented amounts of experimental, simulated, and observational data and through this understanding to address problems previously deemed intractable or beyond imagination"* [JMM+06, p. 6].

Visualization can be broken down into two main categories; visualizations that consider data that represent phenomena that have a known physical or conceptual representation (e.g. the

human body, the earth, or a molecular structure) and those that represent phenomena that are abstract and have no known conceptual representation (e.g. DNA sequences or the relationships between people within criminal networks). Card *et al.* define these as scientific visualization and information visualization respectively [CMS99, p. 7]. Note that this categorization is somewhat misleading, as information visualization is widely used to visualize scientific data. Although a more appropriate categorization would be concrete and abstract visualization, we adhere to the common usage as laid out by Card *et al.,* using the generic term visualization to encompass both concepts. Although the research presented in this dissertation does not explore visualization algorithms and visualization interaction techniques explicitly, the visualization of complex scientific data is one of the essential tools that are used by the scientists that this research considers. Thus visualization, and the processes we use to explore complex data, are fundamental to our understanding of how researchers collaborate. We explore some of the related research on frameworks for the exploration of scientific data in more detail in Section 2.3.4.8.

2.2 The Science of Collaboration

We now switch from studying how researchers collaborate to exploring how researchers study collaboration. The process of groups working together has been under extensive study for over a hundred years [PS99, p. 1], and much of this research is relevant to understanding how scientists work together. Scientific collaboration, and in particular artifact-centric scientific collaboration, is a group process. The study of group work is complex and there are many fields of study that must be taken into consideration when studying this collaboration process.

Cognitive psychology informs us on how the mind processes information and the limits of our cognitive processing abilities. The study of human communication informs us that there are many types of communication, that communication can be modelled as a process, and that communication can be both verbal and non-verbal. Social psychology, the study of how individuals interact in groups, suggests that how we communicate ideas and concepts is very complex and that both verbal and non-verbal communication are fundamental to this communication. Interestingly, the study of non-verbal communication, and gesture in particular, has recently experienced a resurgence of research interest over the past two decades. Clearly, the social psychology community's research into how we use gesture has a lot to contribute to our study of how scientists interact with complex digital artifacts.

Over the past 40 years, the CSCW community has built on, and complemented the research discussed above, exploring how technology can be utilized to support group work more effectively. This includes the use of technology to support groups in the same room (collocated group work) as well as groups that are distributed across multiple physical locations (distributed group work). At the same time, research into HCI technologies (such as touch sensitive screens) and techniques (gestural interaction), digital media (high definition video streaming), and advanced networks have converged to create a technological environment that enables new types of collaboration environments.

In many ways, we are seeing a convergence of both opportunity and need for the support of distributed, artifact-centric, scientific collaboration. Scientific collaboratories are becoming commonplace, and yet their needs are poorly understood. Computational science is producing data at an unprecedented rate, and yet effective artifact-centric collaboration tools are essentially non-existent. Gestural interaction is seeing a resurgence in research interest in the social psychology community, but it is not supported in remote collaboration tools. Touch sensitive devices are becoming ubiquitous (from the phone to wall displays), and yet we have failed to develop compelling group work tools that make use of them. We explore both the opportunity and the issues of these research domains in more detail below.

2.2.1 Communication

The study of communication covers a broad range of types of communication. Huebsch divides up human communication into five main types, interpersonal communication, intrapersonal communication, extrapersonal communication, mass communication, and media communication [Hue89]. We are primarily concerned with interpersonal communication. Huebsch describes interpersonal communication as communication that takes place between two people, presupposing dyadic (two person) interaction where either verbal or non-verbal communication (or both) could be used [Hue89, pg 8]. In particular, non-verbal communication is classified into several types, including general appearance of the person (including attire), facial expressions, paralingual voice characteristics (inflection, resonance, and rhythm), kinesics (human movement), and proximics (space and territory). Kinesics is further divided into several types of

movements, including emblems (the V sign for Victory), adaptors (fidgeting), illustrators (complementing or emphasizing words), gestures (motions with the hands and arms), postures (changes in body position), and regulators (shrugs, head shakes). The research presented here considers both verbal and non-verbal communication, with a focus on how gestures are utilized when interacting with scientific data. We explore both verbal and non-verbal communication in more detail below.

2.2.1.1 Communication Models

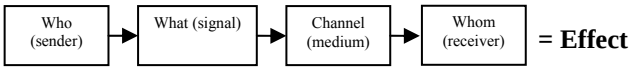


Figure 1: The Lasswell Maxim

In communications research, one of the simplest models of communication is that presented by the Lasswell maxim "Who (says) what (to) whom (in) what channel (with) what effect" [Las48]. Three components of this statement are critical in distributed communication, that of the information or signal being communicated (the what), the medium being used to communicate that information (the channel), and the desired result of the communication (the effect). As discussed above, although the channel used to communicate information is often verbal in nature, a wide range of non-verbal channels (facial expression, gesture, body language) can also be used. From a sensory standpoint, the receiver of the communication receives information using many sensory streams, with the auditory and visual sensory systems the primary mechanisms with which information is processed.

A common extension to the Lasswell maxim is the Shannon and Weaver model [SW49], which differentiates between the sent and the received signals through the possible addition of noise to the signal. In human-to-human communication, the addition of noise may be manifested in a variety of ways, including not understanding a verbal statement, misinterpreting body language, or not seeing a gesture that is used for emphasis. In the case of distributed communication, noise may manifest itself as poor quality audio (possibly caused by actual digital noise in the signal), a low fidelity video feed (which does not achieve the desired effect) or the complete loss of a communication channel (due to it not being provided to the remote user).

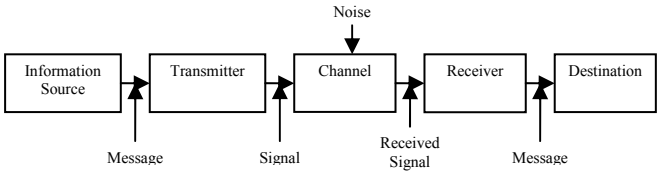


Figure 2: The Shannon and Weaver Communication Model

Berlo's Source/Message/Channel/Receiver (SMCR) model [Ber60] extends the Shannon and Weaver model on several very important dimensions. First, the SMCR model adds social factors that might affect the encoding and decoding of the message. These factors include things such as communication skills, attitudes, knowledge, and the socio-cultural environment. SMCR divides the message up into content, the elements of that content, the treatment applied to those elements to create structure, and the coding of that structure for transmission. SMCR is also one of the first communication models to discuss the channel used to communicate a message, where the channel is represented by sensory channels (seeing, hearing, touching, smelling, taste) rather than communication channels (speech, writing, etc).

Other models consider relevant communication characteristics. Barnlund's Transactional Model [Bar70] incorporates the environment and the context of the communication more completely. Watzlawick and colleagues [WBJ72] point out that when two or more individuals are interacting through any type of communication channel, we need to consider that the absence of behaviour (for example, silence) communicates information to others in the group. One of the earliest models to incorporate two-way interpersonal communication is Schramm's Interactive Model [Sch54]. Note that in this model, each individual plays both the role of an encoder and decoder of information. It is also worth noting that the Schramm model is also one of the first communication models to consider the context (the environment, both personal and physical) of the communication as an important factor in the communication.

2.2.2 Social Psychology

Social psychology is the study of relationships among people, in particular how people work in groups. As such, it is a fundamentally important domain of research in the context of distributed, scientific collaboration. Social psychology is an interdisciplinary

area of study that exists at the intersection of sociology and psychology. Although research in this area occurs at the intersection of these two domains, the two disciplines often take different perspectives. For example, in the study of how language is used in groups, researchers from the cognitive sciences tend to focus on speakers and listeners as individuals while researchers from the social sciences tend to study language as a social process [Cla96, p. 4 and p. 24]. As Clark suggests, it is critical that both perspectives be considered. In this research, there are two main domains of social psychology that we delve into deeper – the study of language use in groups and the study of gesture as a means of communication.

2.2.3 Language

Language is a fundamental part of any communication, and it is therefore necessary to consider it in some detail if we hope to understand how scientists collaborate. By language use, we do not meant the traditional linguistic approach to understanding language (morphology, syntax, phonetics, and semantics), but rather how language is used from a social psychology perspective. In particular, we use Herbert Clark's views on language use to provide an overview of language use in communication. Clark, in his book *Using Language*, states the main thesis of the book as follows: *"Language use is really a form of* ***joint action.*** *A joint action is one that is carried out by an ensemble of people acting in coordination with each other."* (Clark's emphasis) [Cla96, p. 3]. From Clark's point of view, language use is about communication among people. Clark stresses that language involves both individual (psychology) and social (sociology) processes, and states that *"We cannot hope to understand language use without viewing it as joint actions built on individual actions. The challenge is to explain how all these actions work."* Clark lists six propositions about language use:

1. Language fundamentally is used for social purposes;
2. Language use is a species of joint action;
3. Language use always involves speaker's meaning and addressee's understanding;
4. The basic setting for language use is face-to-face conversation;
5. Language use often has more than one layer of activity;
6. The study of language use is both a cognitive and a social process.

Two of these propositions are of particular relevance to this research. Recall that most communication models (Section 2.2.1.1) incorporate an encoding process (encoding the speaker's meaning) and a decoding process (decoding to create the addressee's understanding). This concept (Proposition 3) is important because the vast amount of literature on non-verbal communication has focussed on the encoding process to the detriment of the study of the decoding process [BC06].

Also of importance is Clark's proposition that face-to-face conversation is the basic setting for language. Clark cites Fillmore's view that *"… the language of face-to-face conversation is the basic and primary use of language, all others being best described in terms of their manner of deviation from that base"* and Fillmore then goes on to say *"I assume that this position is neither particularly controversial nor in need of explanation"* [Cla96, p. 8]. Clark lists the following set of characteristics of face-to-face conversation that differentiate it from other types:

- Co-presence – participants share the same space;
- Visibility – participants see each other;
- Audibility – participants hear each other;
- Instantaneity – participants perceive each others actions with no perceptible delay;
- Evanescence – the medium fades quickly;
- Recordlessness – actions leave no record or artifact;
- Simultaneity – participants can produce and receive at once;
- Extemporality – Formulate and execute actions in real time;
- Self-determination – Participants determine what actions to take and when;
- Self-expression – The participants take actions as themselves.

The first four characteristics demonstrate the immediacy of face-to-face communication, the next three characteristics consider how the communication medium impacts the communication, and the last three characteristics consider how the communication is controlled. When considering distributed collaboration, it is important to consider each of the characteristics of face-to-face communication to determine the impact that distance has on the communication that takes place.

Clark also breaks down language use into three fundamental concepts: joint activities, joint actions, and common ground. Joint activities are activities that are goal defined

social events that are bounded by participants, setting, and cultural activity (e.g. teaching, shopping, a dinner party, or a work meeting). Joint activities are carried out by performing joint actions which are coordinated between two or more people. Actions coordinate both what the participants intend to do and how they do it. Finally, common ground between two or more people is *"… the sum of their mutual, common, or joint knowledge, beliefs, and suppositions"* [Cla96, p. 93]. Common ground is of particular relevance to this research, as common ground has been shown to be important for artifact-centric collaboration as well as language use [BG07].

Clark's social psychology approach to language use also differs from the linguistic view of language in that Clark views non-verbal communication as a fundamental part of language. Clark defines signals as *"… the acts by which one person means something for another…"* [Cla96, p. 155]. Rather than hold to the traditional view of signals as linguistic entities (speech sounds, words, sentences), Clark stresses the importance of non-linguistic acts in language use, suggesting that signals are not exclusively linguistic or non-linguistic but instead that most signals are*"… composite signals, the artful fusion of two or more methods of signalling"* [Cla96, p. 156]. Clark goes one step further, stating that non-linguistic methods *"… are part and parcel of most signals that are usually classified as 'linguistic'"*. In many ways, Clark's book on language use is a criticism of the research community's treatment of language, stating that *"Ignoring non-linguistic methods has distorted people's picture of language use, and it is important to put that picture right"* [Cla96, p. 156].

Language in the form of the sentence has long been studied by linguists and philosophers. From a linguistics point of view, the syntax and semantics of language imply that the meaning of a sentence is the composition of its parts. Like Clark, we assume this to be true [Cla96, p. 161]. Clark points out that utterances are not sentences. For example, from a linguistic standpoint, the sentence "I like that one in the corner" can be completely analyzed in terms of its syntax and semantics. From a social psychology perspective, the sentence does not take on communicative meaning until it is uttered by an individual (which gives communicative meaning to the word "I") in a specific physical context (which gives meaning to the phrase "in the corner"). In addition, the phrase "that one" does not take on meaning until an individual makes the utterance at the

same time as performing a non-verbal action that disambiguates the object to which the utterance refers. Clark views the communicative event as being the composition of all of these communicative actions (the utterance of the sentence, the gesture, and the context).

2.2.4 Gesture

"As to the hands, without the aid of which all delivery would be deficient and weak, it can scarcely be told of what a variety of motions they are susceptible, since they almost equal in expression the powers of language itself, for other parts of the body assist the speaker, but these, I may almost say, speak themselves"

Quintilian, *Institutio oratoria*, XI, III, 85 (35-100 CE).

Gesture has been studied since classical antiquity, when Greek and Roman philosophers studied the use of rhetoric (the art of speaking effectively) as part of the process of public oratory. The most famous (and today most complete) classical treatises on rhetoric, are Aristotle's *Rhetoric* (384-322 BCE), Cicero's *De Oratore* (106-43 BCE), and Quintilian's *Institutio oratoria* (35-100 CE). Aristotle discouraged the use of theatrical non-verbal techniques (delivery) such as gesture and tone of voice for persuasion, believing that argument should be based on fact and fact alone, but did recognize that *"…since the whole business of Rhetoric is to influence opinion, we must pay attention to it [delivery], not as being right, but necessary"* [Aristotle's *Rhetoric,* Book 3.1]. Later treatise by Cicero and Quintilian regard theatrical delivery mechanisms such as gesture, facial expression, and tone of voice important to the development of an orator.

It is difficult to separate the use of language from physical action. When individuals are in a face-to-face communication, they communicate information through bodily actions, including information about intentions, interests, feelings, and ideas. Body position, eye gaze, and gesture all provide information to those who are part of such a communication. As Clark suggests [Cla96], many of these actions are a fundamental part of communication with actions complementing, supplementing, or substituting for words during discourse.

Clark's model of language use is very closely aligned with the views of leading social psychology researchers that study gesture as part of communication. The two seminal books in gesture research over the last 20 years, McNeill's *Hand and Mind: What*

Gestures Reveal about Thought [McN92] and Kendon's *Gesture: Visible Action as Utterance* [Ken04] stress that utterances consist of both linguistic and non-linguistic components. Kendon defines gesture to be *"A visible action when it is used as an utterance or as a part of an utterance"*. Kendon then defines an utterance as *"… any ensemble of actions that counts for others as an attempt by the actor to give information of some sort"* or any *"…unit of activity that is treated by those co-present as a communicative 'move', 'turn', or 'contribution'"* [Ken04, p.7]. Both Kendon and McNeill put forward the concept that *"… gestures are an integral part of language, as much as words, phrases, and sentences – gesture and language are one system"* [McN92, p. 2].

The perspective that gesture and speech are produced by a single cognitive system, and even the question about whether gesture is communicative, is relatively new. Early seminal work in this area was performed by both McNeill and Kendon, with initial research being carried out in the eighties [McN85, Ken80]. Although the theory became established in the nineties, with books by Clark [Cla96] and McNeill [McN92] solidifying both the theory and its application, it is still somewhat controversial today [BKJ+02]. This debate revolves around determining whether gesture plays a communicative role (communicates information in and of itself) or a facilitative role (help formulate and support verbal communication). Bavelas *et al.* suggest that it is this controversy that has sparked a renewed level of interest in gesture research [BKJ+02].

Speech, when used as part of an utterance, is relatively easily understood. Gesture is much more difficult. How do we classify a gesture as part of an utterance? What do we use in our analysis and what do we ignore? McNeill's definition of gesture complements Kendon's stating that gestures are *"… the movements of the hands and arms that we see when people talk. Sometimes the movements are extensive, other times minimal, but movements there usually are"* [McN92, p. 1]. Based on the above definition of gesture, McNeill categorizes gesture into iconics (direct pictorial), metaphorics (abstract pictorial), beats (rhythmic), cohesives (linking), and deictics (pointing) [McN92, p. 12].

Since we are primarily concerned with artifact-centric collaboration, we focus on deictic gestures. Deictic gestures are those that point at either physical objects/events in the concrete world or abstract representations of objects in the non-existent narrative

world. It should be noted that McNeill, because of his interest in gesture use in narrative, is primarily concerned with deictic gestures that refer to non-existent entities in the narrative world while we are primarily concerned with deictic gestures at digital artifacts. In the extensive overview presented in *Gesture: Visible Action as Utterance*, Kendon states that *"… gesture of pointing, or deictic gestures, have been recognized as a separate class by almost all students [researchers] of gesture we have reviewed and it has always been understood that such gestures play a fundamental role in establishing how an utterance is to be understood"*. Interestingly, he also points out that *"There are very few studies, however, which have examined the way in which pointing is done"* [Ken04, p. 199].

Bavelas *et al.* provide an excellent review of much of the recent research on the use of conversational hand gestures in face-to-face communication [BG07]. The authors emphasize three of the fundamental requirements laid out by Clark for face-to-face conversation, that of visibility, audibility, and instantaneity (see Section 2.2.3) [BGS08]. They also stress the importance of experimental research, raising the issue that much of the current research in this area considers individuals as the unit of measure (as opposed to dyads or groups). Another issue they identify is that researchers have been limited to measuring the level of communication indirectly (most research counts gesture movements) and are not able to directly measure the level of meaning communicated. Recent research has shown that both of these limitations can be overcome, making this an exciting time for gesture research [BG07]. We now consider some of this research.

2.2.4.1 Gesture, Understanding, and Memory

One important challenge in the gesture literature is the ability to measure the impact of gesture on meaning and understanding. For example, Clark and Krych show that instructor/builder dyads are able to complete building tasks using Lego blocks faster when the instructor can see the builder's workspace [CK04]. Through a detailed analysis of gesture/utterance interactions, they are able to show that dyads correct errors and understand faster when the workspace is visible. It should be pointed out that studies that show an impact on understanding (not just on task performance) are relatively rare. Church *et al.* used speech only and speech+gesture videos to test subjects' processing and recall of gesture [CRG07]. They found that participants were more likely to recall items

in the speech+gesture condition than in the speech only condition. They also found that gesture enhanced recollection in both conditions. These results suggest that gesture helps to strengthen memory recall about speech and that it may play a role in helping with the durability of those memories.

2.2.4.2 Gesture and Maintenance of Conversation

It has been shown that gestures are used to aid in the maintenance of conversation. In particular, gesture use is significantly reduced when a speaker is talking alone, talking to an addressee who could not see them [BCL+92], and talking to a visible addressee but in alternating monologue rather than dialogue [BCC+95]. Similarly, participants were found to use more gestures when they believed someone was watching them re-tell a story [MKM+09]. Interestingly, this study also showed a significant correlation between the gesture frequency and whether the participants thought the listener was human or not [MKM+09]. This result is significant, as it suggests that not only do we use gesture more when we know we are talking to another person, but we also recognize that gesture is used more often when talking person-to-person. The fact that this study considers both the encoding (how gestures are used to communicate) and decoding (how we process gesture communication) makes it relatively unique, as studies that consider both processes are extremely rare.

2.2.4.3 Gesture and Shared Space

Common Ground has been shown to be important in conversational dialogue [Cla96]. Gestures have also been shown to play a similar role. Furyuma used an instructor/builder task involving origami, where the instructor was asked to teach a builder to make an origami figure without using origami paper [Fur00]. Builders often gestured in the instructor's work space, referring to the space in which gestures were just made by the instructor, effectively creating a virtual workspace that did not physically exist but helped to coordinate the task. Ozyurek showed that the location of the shared space in which dyads worked influenced the direction and orientation of their gestures [Ozy02]. The study provides evidence that gestures are indeed communicative (they are targeted at addressees), that dyads share a common space when communicating, and that gestures (but not speech) adapt to the spatial orientation of that shared space.

2.2.4.4 Gesture, Coordination, and Attention

Bangerter investigated pointing gestures for coordination with an experiment that used dyads that talked and gestured to identify arrays of visible targets (photos of faces) [Ban04]. Ambiguity of pointing was operationalized by changing the distance to the photos (arm length, arm length + 25cm to 100cm). Visible gesture was controlled by some pairs being able to see each other while others pairs could not. The study showed that visible pairs (gesture could be seen) used more points with deixis (using pronouns like this, that, here, there) and fewer words as targets got closer. Post-hoc analyses revealed that pairs used more points with deixis at arm-length than other distances. Hidden pairs used the same number of words independent of distance.

Bangerter also suggests that pointing gestures focused attention by directing gaze to the target region. This argument is based on three facts in the arm's length condition. First, the identification task is completed successfully in the arm's length condition. Second, there is an increase in the frequency of deixis pointing. Third, there is a decrease in the frequency of descriptions that refer to location and descriptions that refer to identifiable features. The author argues that this implies that deixis pointing is increasingly being used to focus attention on the targets as the targets get closer. The author also suggests that the decrease in use of deixis pointing gesture and the increase in description as the targets get farther away is likely the result of the gesture being more ambiguous with distance.

Bangerter and Chevalley extend these results with a further study that increases the density of the objects that subjects are required to identify, thereby increasing the ambiguity of the pointing gesture [BC07]. This study shows that subjects use deictic pointing gestures to both indicate referents (indicated by replacing feature descriptions) as well as to draw attention to regions (indicated by replacing location descriptions). Note that one of the key flaws of these two studies is that they assume that the reduction of feature and location descriptions, combined with an increase in pointing gestures and the successful completion of the task, are effective measures of attention being drawn to objects and regions. Although this argument is a reasonable one, the study only indirectly measures the attention paid to the referent artifact that results from the pointing gesture.

2.2.4.5 Gesture Summary

Gesture has been studied since antiquity, and has long been viewed as an important tool for communication. Although somewhat embroiled in controversy, many social psychology researchers believe that gesture and thought are part of the same cognitive process and that gesture, in and of itself, communicates meaning during conversation. It is these communicative gestures that we focus on in this research. Since we are primarily interested in collaboration around scientific data, pointing and deictic gestures are of particular interest. In the sections above, we summarize much of the recent research in gesture, showing that gesture helps in understanding and recall, that gesture helps in the maintenance of conversation, that gesture is used to create a shared work space, and that gesture is important in drawing attention to physical objects. In the remainder of this dissertation, we build on this extensive body of research in our exploration of distributed, artifact-centric, scientific collaboration.

2.2.5 Cognitive Psychology

We focus on two aspects of cognitive psychology that are relevant to this research. Since our distributed collaboration tools present complex information using a variety of modalities (aural and visual information), it is important that we consider how our cognitive system processes information. We consider three areas: working memory, cognitive load, and attention.

2.2.5.1 Working Memory and Cognitive Load

There is extensive evidence from the cognitive psychology literature that humans have a very limited working memory (commonly called short-term memory) while having a large, long-term memory. In 1956, Miller coined the phrase “the magical number seven, plus or minus two,” which describes the number of items that humans can store in working memory [Mil56].

In 1932, Bartlett suggested that humans use schema (conceptual structures that help in reasoning) to represent complex concepts [Bar32]. Such schema are used to hold concepts in long-term memory and assist in the transfer of information from long-term memory into limited working memory. Tindall-Ford *et al.* suggest that schemas operate on a continuum between automatic and conscious control. When concepts are being learned, they require conscious effort as part of the learning process. As the concepts are

learned, they are cast into schema which allow for the automatic processing of that concept [TCS97]. While conscious processing of information and schema requires a relatively large cognitive effort, automatic processing of schema in working memory requires relatively little. Tindall-Ford *et al.* present the following example. Recognition of the letter "a" requires conscious effort and thought when a child is first learning to read. As the child progresses, recognition of the letter "a" becomes automatic because the concept is cast as a schema in long-term memory. Thus, recognition of the "a" moves from a conscious effort that requires significant cognitive load to automatic recognition that requires limited cognitive load. Over time, such schemas develop that help us to automatically process words and sentences. Automated schemas are also believed to be critical to the transfer of information from long-term memory to working memory. If schemas did not exist, the working memory would be overwhelmed when any complex problem was considered [TCS97].

It is important to consider the cognitive processing implications of how information is being presented. Tindall-Ford *et al.* break down the sources of cognitive load into two types, intrinsic and extraneous cognitive load. Extraneous cognitive load is generated by the way information is presented. For example, having spatially disjointed information that needs to be integrated for understanding has been shown to result in high cognitive loads [TCS97]. Thus, when presenting information about a shared artifact in a distributed collaboration environment, it is important to have a coherent work space where all information relevant to an artifact is spatially collocated. This is an example of how collaboration system design might impact extraneous cognitive load.

Intrinsic cognitive load is determined by both the intellectual complexity of the topic and the expertise of the participant. That is, if the topic requires the interaction of many elements (where each element requires the use of working memory) and attaining understanding requires the integration of all the elements, then intrinsic cognitive load will be high. At the same time, if an expert is a participant in the communication and he/she has incorporated all of the elements into a single automatic schema, then he/she will be able to process the same information with very little cognitive load. Thus, when developing tools for distance collaboration, it is important to consider the cognitive

processing implication of both how information is presented and the diversity of the audience to which the information will be presented.

Unfortunately, only one of these two factors can be influenced by distributed collaboration tool builders. We can carefully design to minimize extraneous cognitive load (by presenting information effectively) but intrinsic cognitive load can only be changed through the creation of new schema for understanding. That is, distributed collaboration is "replacing" a familiar collaboration modality (face-to-face collaboration) with an unfamiliar modality for which we have little or no cognitive schema (a given remote collaboration scenario). Thus, when designing new collaboration systems, it is important to consider how to leverage existing cognitive schema that might already exist (build on the familiar) as well as facilitate the acquisition of new schema (present a collaboration environment that is consistent in the mechanisms by which it presents information so that it can be learned) within the new collaboration environment.

2.2.5.2 Attention

Attention, from a cognitive psychology perspective, is the process of selectively focussing on one aspect of the environment while ignoring others. Knudson describes our cognitive processing of the world around us as follows: *"To behave adaptively in a complex world, an animal must select, from the wealth of information available to it, the information that is most relevant at any point in time. This information is then evaluated in working memory, where it can be analyzed in detail, decisions about that information can be made, and plans for actions can be elaborated. The mechanisms of attention are responsible for selecting the information that gains access to working memory"* [Knu07]. The study of attention is an extensive area of research, and we are interested in attention from two perspectives. First, in order to understand how researchers process information about digital artifacts, we need to understand how attention is drawn to those artifacts during the communication process. Second, we need to understand how the directing of attention impacts cognitive processing and understanding.

Attention is often divided into two processes, overt attention and covert attention. Overt attention is the process of physically directing our sensory system (e.g. our visual system) towards a stimulus (e.g. looking at an object) while covert attention is the process of mentally directing attention (e.g. mentally focussing on a single part of a computer

screen while looking at another). Although it has been shown that covert shifts in attention do not necessarily result in an overt shift of the eyes, it has been shown that spatial attention and the eyes often move about the environment in tandem [HK03].

This is of critical importance to this research because we use gaze fixation as a measure of attention (see Chapter 8). Although there is not necessarily a one-to-one relationship between eye gaze and attention, eye gaze is a key indicator of attention. In fact, gaze control is one of the seven key components in Knudson's "Fundamental Components of Attention", playing a role in both the processes of competitive selection (determining what information makes it into working memory) and sensitivity control (regulation of the relative signal strengths of the different information channels that compete for access to working memory) [Knu07].

The tight coupling between attention and eye gaze is also clear. In tasks where an individual is being asked to manipulate a physical object (requiring attention), it has been shown that gaze fixates on the object on average 250 ms after the end of the word that uniquely identifies the object [TSE+95]. When there is ambiguity in the environment, fixations are spread between multiple objects. When the object is no longer ambiguous, fixations rapidly focus on the target object. Indeed, gesture itself is impacted by this ambiguity as well. For example, when subjects are presented with two artifacts on a computer screen and asked to point at one of the artifacts as soon as the artifact is clearly identified through an aural description, mouse trajectory tends to be attracted to both artifacts (often the space between them) until the aural description disambiguates the artifact of interest. Once the artifact is no longer ambiguous, the pointing gesture rapidly converges on the correct artifact [SD06].

2.3 Computer Supported Collaborative Work (CSCW)

In Section 2.1, we explored how researchers collaborate. In Section 2.2 we explored how researchers study collaboration. In this section, we explore how computer scientists study the domain of computer mediated collaboration. In the oft-cited "Baecker Book" [Bae93], CSCW is defined as *"… computer-assisted coordinated activity such as problem solving and communication carried out by a group of collaborating individuals"*. CSCW concentrates on technology mediated human-to-human communication rather than human-computer interaction. This human-to-human focus has

made it clear to the CSCW community that to be successful it is necessary to consider both behavioural science research (such as the psychology and sociology of groups, conversational analysis, and linguistics) as well as technological research (human-computer interaction, networking and communications, user interfaces, audio and video technology, and intelligent systems) [Bae93].

Most CSCW research considers one of two types of systems; those that support collocated collaborative work and those that support distant collaborative work. Collocated collaborative work is computer-mediated collaboration that takes place between individuals in a single physical location. Distant collaborative work is computer-mediated collaboration that takes place between individuals who are physically distant to one another.

2.3.1 Collocated Collaboration

Collocated collaboration makes use of a wealth of subtle cues that we interpret as part of the communication. Not only do we speak to each other, but we also gesture at objects, use props, and "talk with our hands". In addition, very subtle mannerisms such as facial expression and body language play an important role in our understanding of each other. Our physical environment is a very rich one that includes information from a variety of sources, which we often use simultaneously and switch between quickly and easily [FJH+00]. It is these aspects of communication that we must take advantage of and facilitate when exploring technology-mediated collocated collaboration. It is also clear that when considering distributed collaboration, the issues that are pertinent to collocated collaboration are also highly relevant to distributed collaboration. That is, how can we maintain the high level of interaction that is present in collocated collaboration when that collaboration takes place at a distance? We therefore need to carefully explore collocated collaboration as a precursor to trying to understand distributed collaboration.

2.3.1.1 Collocated Object-centric Collaboration

Before exploring either artifact-centric collaboration or distributed collaboration in detail, it is helpful to explore the area of collocated, object-centric collaboration. We define object-centric collaboration as collaboration that focuses on real, physical objects (as opposed to digital artifacts that exist on a computer screen). This is critical to

developing tools for distributed artifact-centric collaboration, as how we work collocated with physical objects defines the human interactions with those objects that we need to replicate in an artifact-centric collaboration system. In particular, the naturalistic study of how we collaborate with physical objects (i.e. studying collaboration in a natural setting), with the goal of developing better tools for technology mediated collaboration, is fundamental to advancing both collocated and distributed collaboration.

Several key studies in this area have had long-standing impacts on the CSCW community. In particular, the task of collocated design has been explored extensively. One important early effort in this area is the work of Tang *et al.* on studying how design teams work together in a naturalistic, collocated environment [TL88, Tan89]. Their study shows the importance of gestures in communicating information about physical objects. They observe that up to 35% of the gestures used either refer to objects in the workspace or enact simulations of those objects. Another important research thread is the series of studies that is summarized by Bekker, Olson, and Olson [BOO95]. Like Tang, Bekker *et al.* also show that gesture is a key component of collocated design teams, with up to 14 gestures used per minute. They summarize their observations as follows:

- Many gestures are brief;
- Gestures are synchronized with speech;
- Gestures often occur in sequences;
- Gesture is often interleaved with other activities;
- Participants move around while gesturing;
- Gestures have complex 2D and 3D trajectories;
- Gestures occur in relation to the spatial relationship of people and objects;
- Gestures often refer to imaginary objects; and
- Gesture can refer to gestures in the past.

Although these observations are specific to the task of design, it appears that for tasks that are object-centric, gestural interaction is a key component of the communication. Such findings are not surprising, as they reflect the findings of the social psychology and gesture research communities (see Section 2.2.4).

2.3.1.2 Technology Mediated Collocated Collaboration

Co-located collaboration that is facilitated by technology is becoming an increasingly active area of research, but until recently this was not the case [SBD99]. Indeed, the extensive collection of pre-1993 CSCW research presented in Baecker's CSCW book [Bae93] has only one of thirteen chapters on collocated collaboration while the others are on distributed collaboration. If non-technology mediated collaboration works so naturally, one might ask why then explore technology-mediated collocated collaboration? The principal reason is that much of our current environment (work and social) is digital in form, and this trend is only increasing. More often than not, when technology is brought into the equation, communication breaks down. Communication becomes a one-way street (the PowerPoint mentality) and "brain-storming" becomes "brain-numbing" as people fight with technology to communicate and collaborate. Collocated collaboration research explores how to extend our natural ability to communicate with others into the digital realm, attempting to bring digital technologies into the communication process in a seamless manner, so that they are part of the communication, not part of the problem.

In 1991, Heath and Luff stated that *"Despite technical advances in CSCW over the past few years we still have relatively little understanding of the organization of collaborative activity in real world, technologically supported, work environments"* [HL91]. That is, most research either studies non-technologically sophisticated collaboration in the "real world" or technologically sophisticated collaboration in the laboratory. For the most part, this is still the case today, with the majority of studies of advanced collocated and distributed collaboration taking place in the laboratory and not the regular work environment. As a result, studies that explore naturalistic work environments, and in particular those work environments that are technologically sophisticated, are key to informing both technology mediated collocated and distributed collaboration research.

Naturalistic techniques have been used to study the collaboration behaviour of a number of technically sophisticated communities in their real-world work environments. This includes Heath and Luff's study of the Line Control Rooms of the London Underground [HL91], as well as their study of collaboration in a doctor's office and in an architectural firm [LHG92]. Of particular interest is Heath and Luff's study of the

London Underground Control Rooms. This study is interesting in that it is a naturalistic study of a real-world workplace where the workplace utilizes advanced technologies to support the collaboration. In essence, they attempt to describe *"… the details of communicative and collaborative work in a real-world environment which incorporates technology similar to that being developed in the field of CSCW"* [HL91]. Such studies that are both naturalistic *AND* study an advanced collaboration technology are rare in artifact-centric collaboration. The main reason for this fact is that technologically advanced, artifact-centric collaboration environments are rare in the traditional workplace. Thus, studies typically explore either advanced technologies in the laboratory *OR* naturalistic studies in the non-technical work place. Rarely is it possible to consider both.

Technology mediated collocated collaboration research has focused on moving away from the one-user/one-computer model that is so prevalent in computer technology today to one of supporting multiple users in a wide range of technology environments. These include the extension of the desktop to support multiple users [SBD99, BF91], digital whiteboards and "SmartRooms" [FJH+00, SBD99, Rek98, PMM+93], and digital tabletops [Wel93, DL01]. All of these systems attempt to take advantage of our natural communication abilities and extend these into the digital environment.

This is of particular importance when one considers the domain of collaboration in the computational sciences. With the increasing size of scientific data sets, and the resultant reliance on the computer to explore that data, how can computer technologies help researchers collaborate effectively? There are three main technology aspects that one needs to consider when creating such a technology environment: image display (how imagery is projected onto the displays in the environment), device interaction (how the user interacts with the display devices), and user interaction (how we interact with software).

Figure 3: Examples of SmartRoom environments at Simon Fraser University.

Image Display

To create a collocated digital environment, it is necessary to provide computer-generated imagery on one or more display surfaces. Modern technologies such as inexpensive digital projectors and LCD/plasma flat panel displays make it possible to create inexpensive multi-screen display environments. Front-projected display systems [PIH01, SLV+02] use standard digital projectors to display imagery on a screen surface. Rear-projected display systems [CFH97, UI97] use a light transmitting projection material as the display surface, illuminating the projection surface from behind. Imbedded displays [SPM+02, SWS+02] make use of flat panel display technologies (LCD, plasma) to provide a display surface mounted on the wall or embedded in a table.

Many advanced collaboration research prototypes make use of multiple display surfaces in a single room, allowing users of the room a high degree of flexibility in how they present information. Multi-screen meeting rooms utilize a range of technologies, including a small number of projectors or flat panel screens to create multiple display surfaces [CZ09][CS07], a large number of tiled projectors or LCD panels to create a single, high resolution display surface [BGM+07][LRJ+06], and mixed wall and tabletop display surfaces [SGH+99][SPM+02][FJH+00]. It is important to note that although there has not been a significant amount of research that considers the design differences between wall and tabletop displays, it has been shown that the perception of visible elements differs on horizontal and vertical displays and therefore design criteria for such displays also differ [WSF+07]. Images of two such environments, as used in the studies presented in this dissertation, are shown in Figure 3.

Touch Screen Interaction Technologies

Although touch screen interaction systems have been available for some time [SLV+02, SPM+02, SWS+02, Smart], large scale touch screen systems are still relatively rare in the commodity world. Interestingly, touch screen systems are rapidly becoming ubiquitous on the small scale, primarily driven by the smart-phone market. Most large scale commodity touch screen systems (e.g. Smartboards [Smart]) simply map touch screen interaction to mouse interaction. For example, touching the screen and dragging your finger around the screen moves the mouse pointer. Tapping the screen multiple times replaces a double click with the mouse button. In many cases, touch screen

interaction is hampered by the affordances of today's software, which is designed to work with a single user using a single mouse. Thus, although the touch-screen devices themselves support quite sophisticated and complex interactions, such interactions rarely map well to today's applications.

Our naturally dexterous ability to use both hands and/or multiple fingers can be taken advantage of in a touch screen interaction environment. Such a multi-touch capability has been shown to be a very natural one and when users are given the opportunity, they regularly perform multi-touch operations [BM86]. Researchers have explored a number of novel user interface techniques [KFB97][TCG+06] using such interaction technologies. Multi-user interaction is also critical to collocated collaboration. Early work in multi-user interaction concentrated on extending standard applications such as the text editor [BF91] or a whiteboard [PMM+93] to a multi-user environment using multiple pointing devices. Many of the interaction technologies listed above support multi-user capabilities, and many of the more recent systems support multi-user applications [SLV+02, SP98, SPM+02, SWS+02, PIH01]. Unfortunately, hardware limitations still remain a challenge, with multi-touch, multi-user systems primarily existing only in research labs.

User Interface Paradigms

The software user interface is a serious challenge in supporting collocated, multi-user interaction. Almost all traditional software is designed to support the traditional Windows, Icon, Mouse, and Pointer (WIMP) single user interface. Research into transitioning from the WIMP metaphor of software to multi-user groupware, although an active area of research for some time [SBD99][TNG06], has yet to reach maturity. It is still necessary to develop most software and tools from the ground up [TSG+06]. This is exacerbated by the fact that due to a lack of multi-user, multi-touch graphical user interface (GUI) APIs, applications that are developed are likely to only run on the hardware platform on which they were implemented.

In addition to the interaction model described above, the actual user interface on large screen devices, and in particular touch sensitive devices, becomes important. One of the unique characteristics of such environments is the fact that users can orient their physical location around the display surface. This is of particular importance in a tabletop

environment and raises some fairly interesting user interface issues [SGM03][TPI+10]. Which way should the GUI of the application face? Should it change based on who is currently active? Should each user be presented with a different interface? Should each user be presented with a subset of the user interface? A number of researchers are exploring new interface techniques, including rotational user interfaces [FBK+99] and user oriented GUIs [VLS02]. Again, these are experimental systems and primarily exist only in the lab.

2.3.1.3 Collocated Tabletop Collaboration

Tabletop based collaboration is one of the most interesting collocated collaboration technologies because of its unique ability to provide face-to-face interaction across the tabletop. Such interaction is natural in a wide variety of settings, ranging from social interaction over a coffee table in the living room through to management level negotiations and planning at a boardroom table. Studies of interaction in the traditional tabletop environment show that people's interactions are natural, fluid, and animated in these environments [Bly88, SGM03]. Studies also show that collaborators may utilize space differently on tabletop environments, making use of personal, group, and storage territories on the tabletop [SC10]. These territorial behaviours occur in both physical [SC10] and digital [PBN09, TR09, TR10] tabletop environments.

As with any technology-mediated collaboration, success of the technology is measured with respect to two things: how well it supports the natural aspects of the collaboration medium (our natural skills at collaborating around a tabletop) and how well it utilizes and integrates the digital capabilities of the technology into the collaboration. Scott *et al.* [SGM03] provides an excellent overview of the issues involved in performing collocated collaborative work on a digital tabletop. In particular, a set of eight design guidelines is presented for creating an environment that supports natural tabletop interaction supplemented with digital interaction. These guidelines imply that the technology should:

- Support interpersonal interaction;
- Support fluid transition between activities;
- Support transitions between group and personal work;
- Support the use of physical objects;
- Support transitions between tabletop work and external work;

- Provide shared access to digital and physical objects;
- Consider the arrangement of users; and
- Support simultaneous user actions.

Although these design guidelines are targeted at tabletop interaction, many of the guidelines are relevant to other collocated collaboration environments, and in particular those that support touch-sensitive direct interaction with digital artifacts.

2.3.1.4 Collocated Scientific Collaboration

Two relevant research projects that explore collocated scientific collaboration are of particular relevance to this research (initially introduced in Section 2.1.1). Huang *et al.* performed a post-hoc analysis (through interviews with scientific staff who used the system) of the use of the MERBoard system, a large screen collocated collaboration environment designed for the Mars Exploration Rover (MER) mission [HMT06]. The authors describe a number of important outcomes of their analysis:

- that the MERBoard system appeared to support synchronous collaboration between researchers and engineers effectively,
- that collaboration patterns between scientists and engineers changed over time,
- that collaboration tasks changed as the mission progressed (less exploratory work was required), and
- that as less exploratory work was required groups ceased to need the support for shared exploration provided by the MERBoard.

This led the authors to three implications for large interactive displays for supporting group work:

- that large displays are not used for interactive work all the time, and that designers should consider how the displays will be used when not being used interactively,
- that multi-display environments should be designed to be flexible and dynamic so that tasks can be migrated to and from the large displays, and
- that large displays should be designed to support exploratory uses where predefined procedures are unknown.

Wigdor *et al.* performed a participatory design and evaluation of a collocated large screen tabletop and wall display system called WeSpace [WJF+09]. The system used a

participatory design with the designers working with astrophysicists to improve their workflow. Note that although the WeSpace system is a collocated collaboration environment, the group evaluated did not normally work together in a collocated environment (although this was identified as a desirable capability). Thus this is not a study of researchers in a naturalistic environment (as the MERBoard system is). The authors report that the scientists found that the collaboration that WeSpace afforded was valuable to their workflow, that using the system resulted in positive changes to the group's workflow, and that tangible scientific results emerged during the collaboration process. Although the author's state that one of these results would not have been achieved without the WeSpace system, it is unclear from the dialogue quoted how this is substantiated (was it a statement made by one of the participants or was it established as part of the analysis). As with the MERBoard, WeSpace appears to support multi-user interaction well, with all participants interacting with the environment with different users leading the interaction at different times.

These results are highly relevant to the research presented in this dissertation, as they explicitly consider the collaborative, scientific research process.

2.3.2 Distributed Collaboration

The CSCW literature has extensively explored the many facets of working together over a distance and it is clear that distance does matter in collaboration [OO00]. Video conferencing has promised a revolution in how we communicate since the first introduction of the PicturePhone by Bell Labs in 1956. In 1988, Egido [Egi88] provided an excellent summary of the optimistic projections and failures of video conferencing systems through the 1970s and 1980s. Egido predicted (with understandable caution) that the uptake of video conferencing might increase during the 1990s due to the rapid development of technology and a more technologically perceptive user community. It is hard to make a strong case for this being true in the 1990s, and only recently has video conferencing moved from a niche market to a strong commercial industry. Today, high definition (HD) H323 video conferencing is a commodity purchase, with an entry level price of about $2500 for a hardware codec and a 1280x720 (720p) resolution HD camera. In 2008, the video conferencing market generated yearly sales figures of over $1B [Wainhouse09].

Much of the distributed collaboration research has focused on trying to reproduce what it is like to physically be present in a face-to-face meeting. Given the fact that face-to-face communication is considered the "gold standard", this is not surprising. At the same time, some researchers have challenged this approach [HS92][OTC+02]. Instead, they suggest that collaboration problems should be framed in terms of needs (including needs that are not met in the physical environment) and that media technologies should be developed to meet those needs (potentially making the distributed space more effective than the physical one). The focus on needs in this context is a critically important one. The requirements of the collaboration need to be understood, so that it can be determined how technology could be used to restore the information streams that are lost due to distance. It has been suggested that if we do this well, it may be possible to design collaboration environments that are more effective than the face-to-face environment [HS92].

2.3.2.1 Task Centric Collaboration

It is clear that understanding the task being performed while collaborating is critical to delivering a successful environment. Egido [Egi88] ascribes many of the failures of video conferencing technology to this fact alone. One of the most often cited rationalizations for video conferencing is the reduction and/or replacement of travel that should result from effective use of this technology. Egido's findings indicate that although video conferencing has not been successful in eliminating the need for face-to-face meetings, it has been successful in improving communication by increasing the ease with which collaborators can meet. The results we present in Chapter 6 echo this finding, as we show that technology can be an enabler that supports closer collaboration. It does not necessarily remove the need to travel.

Although research into CSCW has been ongoing for many years, even the most basic of information streams are still relatively poorly understood. The exception to this rule is the use of audio in collaborative environments. Early research in CSCW demonstrated that the addition of an audio stream to other mediums improved communication [Wil77]. Few researchers would argue that audio is not a critical component of technology-mediated communication, and work investigating the audio requirements for specific

tasks has been performed in the past [BS97, WS97, KDK99, BS00] and is continuing today.

Contrary to what one might expect, research results on the importance of video in collaboration have been varied. Some studies have shown that video adds little to the ability of users to complete tasks [Wil77, MAF95]. Other studies have shown that having a shared visual environment does improve communication and the ability to perform specific tasks [KGF02, VOO+99, GSH+93, and NSK+93].

The main reasons for these inconsistencies are perhaps understandable – a wide variety of tasks are considered (assembly of objects, problem solving, brain-storming, and formal meetings), the uses of video are varied ("talking head", shared video contexts, and shared computer screens), and the video is controlled differently (frame rate, resolution, image size, and latency). A good example of this is a series of studies performed at Carnegie Mellon University. In [KMS96], the authors perform a study of a bicycle repair task with a remote expert helper where video and audio are used as collaboration tools. The results show that there is no significant difference in the number of tasks completed or in the time it takes to complete the tasks when video is added as a collaboration tool. The authors suggest that one of the reasons for this surprising result is that the video used did not have high enough fidelity on several dimensions (field of view and stability). In [FKS00], the authors improve on their earlier study with better video technology, a more robust study design, and an added side-by-side control group. This study also failed to show a significant impact of video on the task performance, although it did result in conversation that is more efficient. In a more recent paper, the authors again study the use of video in improving task performance, this time on a puzzle task [KGF02]. This study demonstrates a significant improvement in task performance when video is used. In a study of the use of gesture in a physical robot construction task, Fussell *et al.* also show an increase in performance and communication when gestures were communicated using a touch sensitive input device to send video annotations to the builder of the robot but that using a cursor pointer did not increase performance [FSY+04]. What does this tell us? That video is useful in some tasks, but when and how it is useful varies significantly depending on the task being undertaken and the technologies being used to communicate information. In contrasting the studies where video did [KGF02, FSY+04] and did not

[FKS00] improve task performance, what caused these different results? Was the difference because the task was different (between the studies in [FKS00] and [KGF02, FSY+04])? Was it because the video was used to assist in performing the task in different ways? Or was it because the characteristics of the video stream itself (fidelity, quality) were different in the two studies? Despite the recent promising results from this group on communicating gesture effectively to increase task performance [FSY+04], it is still not clear why the video helped to improve task performance in some of the studies and not in others.

A number of studies have been performed on the impact of task on the parameters used to deliver media. In particular, Bouch *et al.* [BSD00] have shown that task can drastically alter what users consider the most important aspect of a certain type of media. Not only do the parameters affect the task, but the interactions between media types also can affect task performance. For example, in [KDK99] the quality of the media streams is not altered but the synchronization of the media streams is, resulting in a decrease in task performance when audio and video are not synchronized. Even more problematic is the impact of poor video on the perception of audio. Rimmel *et al.* [RHV98] present a study in which the user's perception of audio quality decreases as video quality degrades, despite the fact that the audio quality remains constant. The same holds true for the impact of audio quality on perceived video quality. The interaction between the many variables in even the simplest of collaboration tasks is surprisingly complex, with the interaction sometimes changing based on the task being performed.

Although task has long been considered a crucial element of group work [McG93], the above overview of research in this area highlights this importance. It is clear that having a deep understanding of the collaboration task that is being undertaken is critically important if we hope to create collaboration tools that will facilitate the success of that collaboration task at a distance. The approach used in this research to navigate this complex problem domain is through the creation of the CoGScience Framework, a new framework for studying distributed, artifact-centric collaboration. This framework is discussed in detail in Chapter 4.

2.3.3 Distributed Artifact-centric Collaboration

Previous research on face-to-face collaboration suggests that our interactions are naturally multimodal [BOO95][OOM95][OFC+03]. People bring "things" to meetings (physical objects such as a paper document and digital artifacts such as a data set) and refer to these objects on a regular basis. In particular, the coupling of deictic statements with gestures is an important component of most face-to-face collaborations (e.g. look at this) [Ken04, BG07, BC00]. In face-to-face meetings that involve the design process, up to 14 gestures per minute have been recorded [BOO95]. The question then arises - what happens to these communication modalities when collaboration is performed at a distance? Indeed, in their 1995 paper, Bekker *et al.* call for a *"…concerted empirical attack on the question of what happens to gestures during design meetings [when the users are not collocated]"* [BOO95]. Despite this call for action, relatively little effort has gone towards exploring the effects of distance on gestural interaction as a modality for distant collaboration.

2.3.3.1 Distributed Object-centric Collaboration

Before exploring artifact-centric collaboration in detail, it is helpful to explore the area of object-centric collaboration. We define distributed object-centric collaboration as distributed collaboration that focuses on real, physical objects (as opposed to digital artifacts that exist on a computer screen). The computer mediates the collaboration by communicating information about the object to remote users, typically using a video feed of the actual object. Research in this area typically focuses on using video in novel ways to provide data about the task being performed instead of, and some times in addition to, data about the person performing the task. Gaver *et al.* developed a system that used multiple video views of remote collaboration spaces and found that users spent most of their time using video to refer to objects relevant to the task instead of video of the people with whom they were collaborating [GSH+93]. Nardi *et al.* found similar results in an appropriately named study, "Turning away from talking heads: the use of video-as-data in neurosurgery" [NSK+93]. They study how neurosurgeons use video in a technologically advanced operating room and discuss the utility of providing video of a single well chosen object (in this case the surgical area) and the range of functions that video performs (coordination, attention, interpretation, and presence).

Kraut *et al.* presented results on a puzzle task in which collaborators (a helper and a worker) used images of the puzzle to decrease task completion times [KGF02]. Fussell *et al.* report that gesture, when transmitted as annotations during a similar robot construction task also decreases task completion time [FSY+04]. The authors note that the same decrease in task completion time does not occur when a mouse pointer is used to communicate such gestures. Kirk *et al.* have shown in addition that there are effects of receiving remote gestures on worker language, with increases in the amount of words used by the workers and an increase in the amount of overlapping exchanges when remote gestures are not visible [KRF07]. Finally, Fussell *et al.*, in one of the rare user studies that explores how collaborators process information during artifact/object centric collaboration, demonstrates that in a robot assembly task workers attend to (measured using eye tracking) certain artifacts (robot, pieces, and worker hands) displayed in a video stream more than other targets (user manual and the helper's face) [FSP03]. The importance of task, and in particular the importance of the object-centric task as critical in defining the correct information streams to provide for a collaboration task, is clear from these studies.

One of the most concentrated efforts in understanding distributed object-centric collaboration has been carried out by Fussell and colleagues [FKS00] [KGF02] [KMS96][FSP03][OFC+03][FSY+04][OSW+06]. This series of papers focuses on collaboration tasks that involve a physical object, such as repairing a bicycle, solving a puzzle, or building a robot (*object*-centric) and are discussed in more detail in Section 2.3.2.1.

In one of these papers, Ou *et al.* point out some of the key differences between object-centric and other types of collaboration [OFC+03]. In particular, they state:

- Gesture support in CSCW is different from gesture support in human-computer interaction (HCI). Gestures in HCI communicate information to a computer while gestures in CSCW communicate information to other people. One cannot overlook this important role change when creating object (and artifact) centric collaborative environments.
- Gestures may play a role as both HCI and HHI (human-human interaction) communication mechanisms.

- There are few theoretical guidelines to direct researchers in the construction of collaborative environments where gestures play either a HHI or a combined HHI/HCI role.

These observations are very astute, as much of the research performed in the area of gestural interaction has occurred in an HCI context as opposed to an HHI context. It is not clear how much of the extensive gesture research that has been carried out in the HCI community can be applied to human-to-human object and artifact-centric collaboration.

2.3.3.2 Distributed Artifact-centric Collaboration

Artifact-centric collaboration is similar to object-centric collaboration in that the focus of the collaboration is about "a thing". We define artifact-centric collaboration as collaboration that focuses on digital objects or *artifacts* (objects represented digitally on a computer) rather than physical objects. One of the key differences in this type of collaboration is that manipulation of the artifact occurs through interaction with the computer (as opposed to interaction with the physical object itself). Like object-centric collaboration, artifact-centric collaboration is naturally multimodal and research indicates that the communication of gestures is required to perform this task effectively [BOO95, OOM95]. Ou implies that the spatial physicality of the object is what makes multimodal interaction with gestures important [OFC+03]. We point out that a similar spatial physicality exists with artifacts as well, the difference being that in artifact-centric collaboration, the physicality is in the digital realm.

It is arguable that artifact-centric collaboration does not require multimodal interaction to the same level as that of object-centric collaboration. As physical collaboration tasks can be very complex (e.g. repairing a bike), technology mediated object-centric collaboration needs to support complex interactions. We believe that this is also true for artifact-centric collaboration. Even for relatively simple digital artifacts like a spreadsheet, complex multi-modal interactions that include aural, visual, and gestural interactions are common. This holds true for the design process [BOO95] and our research presented here supports this in the domain of scientific collaboration (see Chapter 7). In addition, recall that we are considering scientific artifacts that are structurally complex, possibly changing over time (due to the time varying nature of the phenomena they represent), and poorly understood (little a-priori knowledge about the

artifact is available). We hypothesize that scientific exploration of such an artifact is of a similar complexity (if not more complex) to that of most object-centric tasks, and therefore the information that needs to be communicated is likely to be similar in nature and complexity.

Previous research in distributed, artifact-centric collaboration has included a number of techniques to communicate artifact interaction. Basic mechanisms used to communicate gestural information include the use of telepointers [GP02, Cor03, CS05], avatars [BBF+95], and video overlays [TM90, TNG06, TPI+10, TR09, TR10]. Of particular relevance to this work are the systems that stemmed from the seminal HCI and CSCW work of Tang *et al.* and Ishii *et al.* and the VideoWhiteboard and ClearBoard systems. These early systems used overhead (VideoDraw [TM90], TeamWorkstation [Ish90], and ClearBoard [IK93]) and behind screen (VideoWhiteboard [TM91]) video cameras to project imagery that included both data and gesture to remote collaborators. Given the technology limitations of the time, these systems were impressive feats of engineering. Improvements over these systems have been proposed by a number of researchers over the past twenty years, including the Agora system [KYY+99, LKH+09], the Designer's Outpost [EKL+03], the LIDS systems [AMM+03], WSCS-II [MI04], Kirk *et al.*'s Ways of The Hand system [KCR05], VideoArms [TNG06][TPI+10], C-Slate [IAC+07].

The goal of many of these systems is to explore the integration of task space and person space. Buxton uses the following definitions of space [Bux09]:

- Person space: this is the space where one reads the cues about expression, trust and gaze. It is where the voice comes from, and where you look when speaking to someone.
- Task space: this is the space where the work appears. If others can see it, it is shared. If not, it is private. Besides viewing, this is the space where one does things, such as marking up an artifact or creating new artifacts.
- Reference space: this is the space within which the remote party can use body language to reference the work – things like pointing and gesturing. It is also the channel through which one can sense proximity and anticipate intent.

Buxton defines reference space as the space where task space and person space overlap and points out that task space and reference space do not need to be the same. In fact, in

all but the most sophisticated systems, they are either not the same or the richness of the reference space is highly impoverished compared to the equivalent face-to-face space. Buxton uses a comparison of the Hydra system [Sel92] with the VideoWhiteboard system [TM91] to illustrate this issue. In the VideoWhiteboard system, cameras and projectors are used to create shadows of remote users in a shared workspace, thus creating a rich work space and reference space but an impoverished shared personal space. In the Hydra system, the users have good quality shared personal space but an impoverished shared workspace. Buxton describes this comparison as follows:

> *"In both cases one can see the remote participant(s), and the work being done – in both cases on a large rear projection screen. But here the similarities end rather quickly. In [VideoWhiteboard], one can see no details of the remote person's face, such as their eyes or where they are looking. On the other hand, in [Hydra] the only way that people can point or gesture is with a single point, controlled by a mouse or stylus. This restricts them to the gestural vocabulary of a fruit fly*[3]*. What a contrast to [VideoWhiteboard] where one has the full use of both hands and the body to reference aspects of the work through gestures. In addition, the sharpness and contrast of the shadows provide strong cues that help one anticipate what the remote person is about to do, and where"* [Bux09, p. 228].

Unfortunately, even in state-of-the-art artifact-centric collaboration spaces today, almost exactly the same parallels can be drawn. For example, if one compares the VideoArms system of Tang *et al.* [TNG06] to the Halo system from HP [GDM09], the same parallels can be drawn. VideoArms provides a rich shared task space and reference space but an impoverished shared personal space while the Halo system provides a rich shared personal space but an impoverished shared task and reference space. Although there is a recent renewed research interest in attempting to create rich interaction spaces that include personal, task, and reference space [TPI+10], the fundamental capabilities of these systems are not significantly different than those created by Tang *et al.* in the early 1990s.

Although modern systems make important improvements over the ClearBoard and VideoWhiteboard systems of the 1990s, technologies that support a rich task or reference space have failed to make it into the mainstream. Although Halo [GDM09] and similar

[3] This statement summarizes the problem quite succinctly. The gestural vocabulary of a fruitfly - perfect!

systems are commercial systems, as described above their reference space is relatively impoverished. Most systems that have a rich reference space, such as VideoArms [TNG06][TPI+10] remain research prototypes only. The ability to create a coherent, high fidelity, artifact-centric shared workspace beyond the lab remains elusive.

This limitation brings up a critically important issue in the study of such systems. Most of the systems listed above were created as laboratory prototypes and to our knowledge no artifact-centric system that supports a high-fidelity task and reference space has been used or studied beyond the lab. As suggested by Tuddenham *et al.*, there is *"… little empirical work in the area, and in particular that the work practices afforded by remote tabletop interfaces are not well understood"* [TR10, p. 431]. Although some of these systems have been evaluated with user studies, these studies are primarily qualitative in nature [TNG06][TPI+10] [MI04] [IKG92]. They provide little quantitative data that analyzes the performance of the systems in terms of the types or numbers of gestures used or how effective the system is at accomplishing specific tasks. None of these systems were evaluated outside of the laboratory.

One of the key goals of the research presented in this dissertation is to address some of these issues in the context of distributed, artifact-centric, scientific collaboration. In particular, we perform analyses of how SmartRoom technologies are used by a broad research community (Chapter 6) as well as in depth analyses of how these technologies are used by a focused research group (Chapter 7). Both of these analyses are naturalistic, longitudinal studies of how researchers use these advanced technologies in day-to-day use, and provide us with both qualitative and quantitative results on how carefully designed, advanced collaboration rooms are utilized by scientific researchers.

2.3.3.3 Media Spaces

Media Spaces are one of the few collaboration technologies that have been used extensively in production by scientific researchers. A Media Space is a technologically sophisticated collaboration space that links two or more distributed sites. Originally developed at Xerox PARC in the 1980s to link together offices [Stu09, Har09], the system was later extended to link the Palo Alto and Portland labs (1985) and to link offices within the EuroPARC lab (1986) [Har09, p. 11]. Media Spaces were designed to

help scientists within the Xerox organization work together more effectively. They were also the focus of a significant research effort on CSCW within Xerox.

Media Spaces are as much a concept as a technology. From a conceptual perspective, Media Spaces facilitate the creation of place from space. Space is the reality that surrounds us. Such a space becomes a place through the utilization of the environment to facilitate social interactions among individuals. For example, a meeting room is a space that contains chairs, a table, a computer, and a computer display. At one moment, such a meeting room is transformed into a place for holding a meeting, the next a place for holding a birthday party for a colleague, and the next a place for a casual conversation among colleagues. It is the appropriation of space for a social endeavour that gives place meaning to the participants. Bellotti and Dourish describe a media space as *"… providing a wider set of services based around people's different reasons for wanting to be in contact with each other"* and state that the *"… integration between these components is as much a matter of use as of construction"* [BD97]. Media Spaces, unlike most collaboration environments, are typically designed to be task independent, and in fact are often targeted at creating social spaces (digitally connected water coolers or coffee rooms) rather than task specific meeting rooms.

The VideoDraw [TM90], VideoWhiteboard [TM91], and Team Workstation [Ish90] systems discussed in the previous section were a natural extension of Media Spaces, with VideoDraw and VideoWhiteboard originating from the same group at Xerox PARC. Over the past 20 years, Media Spaces have been used in a variety of domains. An excellent summary of research in Media Spaces can be found in the book *"Media Space: 20+ Years of Mediated Life"* [Har09], which includes a chapter on our analysis of how Media Spaces can be used to support distributed, scientific collaboration (Scientific Media Spaces or SMS) [CZ09]. This work is presented in more detail in Chapter 6.

2.3.4 Collaboration Theories, Frameworks, and Taxonomies

As described at the beginning of this section, the study of distributed, artifact-centric scientific collaboration involves a wide range of disciplines, spanning communications, social psychology, psycho-linguistics, cognitive psychology, gesture, human computer interaction, and computer supported collaborative work. A promising approach to integrating the concepts, models, and theories from these diverse domains is the creation

and application of theoretical and conceptual frameworks. Such frameworks attempt to capture key dimensions of group work. In particular, since we are interested in the study of distributed, artifact-centric scientific collaboration, frameworks and taxonomies that capture the parameters and dimensions of this specific application area are of particular interest. We explore such theories, models, frameworks and taxonomies below.

2.3.4.1 Group Work Theories, Frameworks and Taxonomies

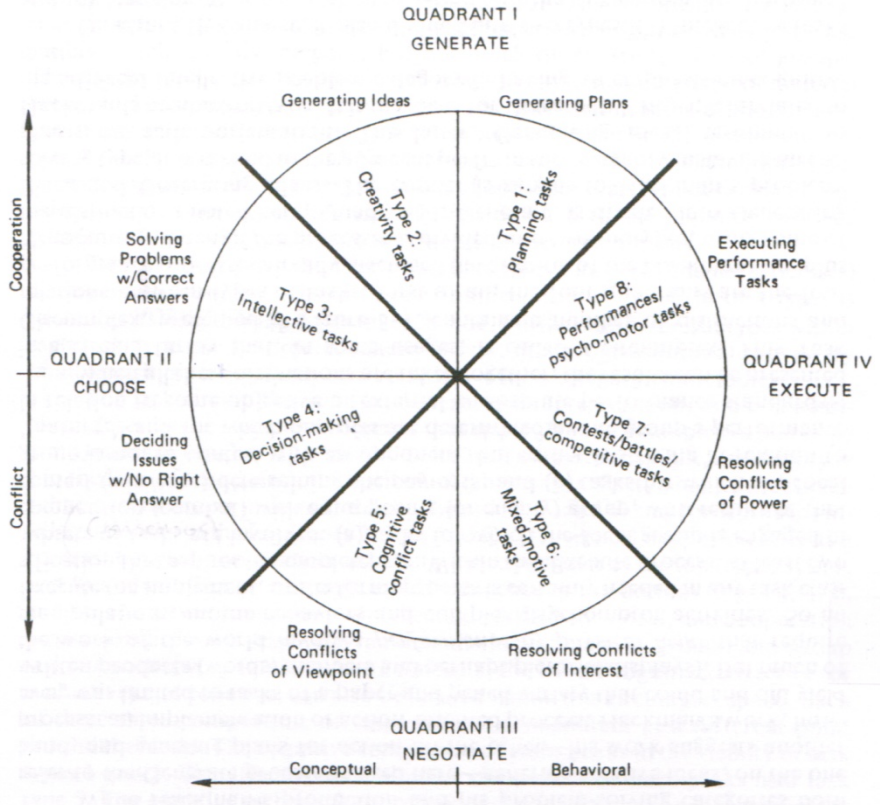


Figure 4: McGrath's Task Typology (Source [McG93])

Perhaps the most fundamental and well known taxonomy for group work is the "Typology of Tasks" proposed by Joseph McGrath in 1984 [McG84, McG93]. McGrath creates a classification scheme that groups tasks into four quadrants and eight distinct categories (see Figure 4). McGrath suggests that the goal of such a classification should be that the categories are a) mutually exclusive, b) collectively exhaustive, and c) logically related to one another. The quadrants represent what the group is to do, and are broken down into generating alternatives, choosing alternatives, negotiation, and execution. Each quadrant is then further broken down depending on the amount of

conflict/cooperation inherent in the tasks or whether the task is conceptual or behavioural in nature. McGrath defines the four quadrants and eight task types as follows.

- *Generating alternatives* tasks include *generating plans* and *generating ideas* (brainstorming);
- *Choosing alternatives* tasks include solving problems with a correct answer (*intellective*) and deciding issues with no right answer (*decision making*);
- *Negotiation* tasks include tasks in which the group is resolving *conflicts of viewpoint* (cognitive) or *conflicts of interest* (mixed-motive, negotiation); and
- *Execution* tasks are either resolving *conflicts of power* (competing for victory) or *performing to meet a standard* (performance).

McGrath's typology is widely used as a fundamental building block for the study of group work, and its impact can be found in many of the other models and frameworks discussed below.

Another important component of the group work domain is understanding how groups interact. Bales' System for the Multiple Level Observation of Groups (SYMLOG) [BC79] is one of the more comprehensive and widely used methods for classifying individual behaviour in a group. In SYMLOG, all actions can be classified as either contributing toward task completion or maintaining inter-relationships between group members. The system uses three main dimensions for measuring group interaction, their dominance/submissiveness, their attitude to relationships in the group, and their attitude toward completion of the task. These are potentially important measures for many group work scenarios.

In 1991, Joseph McGrath produced another seminal paper in group research, entitled "Time, Interaction, and Performance (TIP): A Theory of Groups" [McG91]. McGrath pointed out that much of the empirical foundation of group research theories is based on a limited range of types of *ad hoc* groups examined under experimental conditions. This is not the type of group that we encounter in our day-to-day lives, and therefore many theories are constrained by this relatively narrow scope. The focus of McGrath's TIP theory is to take into account the wider context in which a group exists. This includes taking into consideration:

- The environmental context in which the group is working. Groups rarely work in isolation, and therefore the organization and community in which the work takes place is important.
- The temporal context in which the group is working. The way a group works together at a specific time is impacted by the group's history (work and social interaction in the past) and the group's future (expected interactions in the future).
- The intra-group context of the working group. Individuals are motivated by a range of factors, resulting in work activities that affect the individual (self), other group members, the overall group, and in some cases the organization in which the group is working.

In addition to stressing the wider context in which groups work, the TIP paper also suggests a set of modes in which groups typically function. McGrath suggests group work proceeds in phases, and includes the inception of a project (goal choice), the solution of problems (means choice), the resolution of conflict (policy choice), and execution of the project (goal attainment). McGrath points out that the modes above are similar to the tasks presented above in the "Typology of Tasks" [McG84]. Although there are similarities, it is important to note that TIP is a temporal based theory and explores how groups move through phases over time, while the Task Typology framework [McG84] categorises the characteristics of the tasks that might occur in each of the TIP phases. The recognition that a group's work passes through phases, and that those phases may have very different task characteristics, are important observations.

2.3.4.2 CSCW Theories, Frameworks, and Taxonomies

There are two main types of CSCW theories, frameworks, and taxonomies, those that extend the social psychology and group work to apply to computer mediated collaboration and those that consider how collaboration can be delivered using computer technologies. For example, one of the key implications that McGrath suggests for TIP theory is its ability to help understand the impacts of introducing technology into the group work environment [McG91]. One of the key gaps that exist in the CSCW community is the lack of a framework, model, or theory that bridges the gap between the task domain of group work and the technology domain of computer science at a sufficiently detailed level. Although many frameworks and theories consider both task

and technology, such as Dennis *et al.* Media Synchronicity Theory [DFV08] (see Section 2.3.4.6), it is our belief that both task and technology need to be considered at a more fine grained level. In this section, we explore currently existing CSCW frameworks and theories. In Chapter 4, we present the CoGScience Framework, a new framework that attempts to bridge the gap between task and technology at a level of detail that is currently not available in existing frameworks.

2.3.4.3 The CREW Framework

One of the larger and most comprehensive efforts in the creation of such frameworks is being carried out by the Collaboratory for Research on Electronic Work[4] (CREW) lab at the University of Michigan. The CREW lab has attempted to capture much of the research in this area into usable frameworks for studying group work. The description below is our distillation of two papers [OO97][OO01] into a single framework, which we call the CREW Framework (our naming convention). Many of the concepts span both papers, but the details vary and so we attempt to combine them here. It should therefore be noted that any errors or misinterpretations are solely the responsibility of the author.

The CREW Framework divides group work along several dimensions, including relationships among group members, the work situation, the technology used, and the task [OO01]. Characteristics of the group are individual's characteristics (skill, personality, and motivation) and group composition (homogeneity of abilities, cohesiveness, and trust). Characteristics of the situation are organizational factors (reward structure, work norms, and organizational routines) and particulars of the moment (time the meeting takes place, technologies available, and resources available).

Task characteristics are considered at a level of granularity at which the technology might affect the task. The CREW Framework considers a range of task characteristics, including:

- the nature of the material (abstract ideas or concrete objects);
- the major information processing activity (planning, gathering information, explaining, discussing, or producing a product);
- dependencies among team members (loosely or tightly coupled);

[4] It is worth noting that the CREW lab is very closely related with the Science of Collaboratories (SOC) research group discussed in 2.1, with a large number of researchers belonging to both projects.

- mental resources required (number of constraints and task familiarity); and
- duration and scope (short term to long term).

The CREW Framework also takes the characteristics of the technology into consideration, focusing on technologies that support the conversation and those that support the work object. Factors that support the conversation are what is visible (video clarity, field of view), what is audible (audio clarity, interactivity and echo, spatiality), delay (latency, delay between audio and video), and control over what is sent (passive versus active, control over the channel). Factors that support the work object are control over access (read/write permissions), extent of functionality (editing and navigating with the object), correspondence between views, the ability to locate others, the ability to capture the attention of others, control over turn-taking, support for specific types of work, and ease with which tools can be accessed [OO97]).

In order to understand the many factors listed above, the CREW Framework also proposes a set of measures that attempt to capture the effect of these factors on group process [0097]. These can be thought of as dependent variables that can be measured. Task process measures include depth and breadth of analysis, time spent in various activities, the structure of the work (serial versus parallel) and efficiency. Communication process measures include the pattern of the management of discussion (turn-taking), number of clarifications, non-verbal communication, digressions, and socialization. Measures of interpersonal process include the amount of conflict and cooperation, the mood, and the amount of participation. Measures of outcome include task outcomes, group outcomes, and organizational outcomes. Task outcomes include work product quality (against some criteria). This may be able to be measured directly (how much of the task was completed, task quality) or may require subjective judging. Group outcomes gauge how much the group as a whole supports the outcome. This can be measured as group understanding, group buy-in to the result, or group satisfaction with the result. Organizational outcomes can also be measured in terms of learning (individual and group), willingness to work with the group again, group member loyalty and retention, and individual status changes.

One of the more interesting applications of this framework is given in [OO97], where the authors use the framework for classifying the set of papers that are presented in the

book *Video Mediated Communication* [FSW97]. Through application of the framework, the authors are able to cluster the papers based on the type of group, task, or technology employed, able to point to situations that have been under-tested in empirical work, able to focus on areas where researchers have found apparently contradictory results, and where results are stable and comparable begin to build theory. This analysis demonstrated the value of such a framework. Without a framework to give a common grounding for the research, it is difficult to "make sense of the findings" (the title of the paper). We take a similar approach in the research presented here, applying the CoGScience Framework to the studies discussed in later chapters.

2.3.4.4 The Mechanics of Collaboration

Another relevant effort in this area is Gutwin and Greenberg's work on the *mechanics of collaboration* (MOC) [GG00]. The MOC Framework is particularly focused on collaborations where a shared workspace is involved (artifact or object centric collaboration). The authors define two main aspects of such a collaboration: taskwork and teamwork. Taskwork is the set of actions that are required to accomplish the task successfully. This might include creating artifacts (e.g. documents), organizing artifacts, and exploring the space in which the artifacts exist. Taskwork is the same if an individual or a group is undertaking the task. Teamwork is the set of actions that are required to coordinate the team working together to accomplish the task.

MOC further divides teamwork into the social and mechanical aspects of collaboration. To provide a completely successful collaboration, both of these aspects must be considered. The authors describe a classification for the *mechanics of collaboration* for teamwork and claim that satisfying the mechanics of collaboration is necessary (but not sufficient) to create a successful collaboration system. The MOC Framework has a narrow focus, attempting to capture the mechanisms used to communicate rather than the task specific or social elements of the collaboration. MOC defines seven major activities:

- Explicit communication: Information that is communicated intentionally, including verbal communication, written communication, and deictic gesture.
- Consequential communication: Information that is communicated unintentionally, in particular information that is communicated as artifacts are manipulated by

others and information that is communicated by a person's presence in the workspace (e.g. leaning forward or starting to reach for an artifact).

- Coordination of action: Coordinating turn taking and the process of carrying out ordered actions.
- Planning: Division of labour, dividing up the workspace, and considering plans of action.
- Monitoring: The ability to monitor activity, including awareness of actions in a workspace and the supervision of activities.
- Assistance: Providing help to others to accomplish the task.
- Protection: The ability to access others work and to control access to your own work.

Although the activities that MOC defines are not task specific, Gutwin and Greenberg define groupware usability as *"…the degree to which a groupware system supports the mechanics of collaboration for a particular user and a particular set of tasks"* [GG00]. They suggest that MOC is a useful tool in helping to cast technology mediated group work in the context of a specific task. They also suggest several measures for evaluating the mechanics of collaboration, specifically if groups can perform the mechanics effectively (whether the activity is completed successfully and the number of errors made during completion), efficiently (the resources required to carry out the activity), and satisfactorily (whether the group members are happy with the process and the outcomes).

Gutwin and Greenberg extend the MOC to focus on work space awareness [GG02]. Their framework divides workspace awareness into three components: the information that needs to be captured and sent to remote collaborators (the component elements: who, what, where, when, and how), the mechanisms by which this information is gathered (essentially the MOC), and the way that the information is used in the collaboration (the collaboration processes that the information supports). They suggest that tool designers need to understand how information is used to support workspace awareness to effectively provide distributed collaboration tools where such awareness is required.

2.3.4.5 The ETNA Taxonomy

Just as the MOC focus specifically on the mechanical aspects of teamwork, the Evaluation Taxonomy for Networked Multimedia Applications (ETNA) focuses on

media quality (audio and video media) in the context of collaboration task [WS97, Wat01, MJA+02]. ETNA presents a taxonomy and framework for assessing and evaluating the audio and video media requirements of a distributed collaboration session. Like the CREW Framework, the ETNA taxonomy is task focused, explicitly considering the audio and video communication streams as part of the framework. ETNA uses a decision tree to divide the task domain. At the top level, it considers whether the media is used for telepresence (media represents people) or teledata (media represents other information). It then divides the telepresence branch by considering whether or not the task requires direct user interaction and full attention (foreground tasks) or does not require the full attention of the user (background task). Foreground tasks are further refined if they are interactive tasks or non-interactive. And finally, they refine interactive tasks based on whether the task is a social or negotiation task or whether the task requires problem solving or cognitive processes. ETNA then goes on to consider media characteristics (e.g. video characteristics, such as frame rate required for foreground teledata) and task characteristics (task difficulty, urgency, and emotion) for each of the "leaf nodes" of the task decision tree.

Lastly, ETNA considers the characteristics of the participants in the group (heterogeneity, age, and experience) and the situation (distribution of the users and the physical environment). The ETNA authors suggest thinking of the taxonomy represented as a cube, with the three dimensions representing task, user, and situation [MJA+00]. Unlike most frameworks, the ETNA taxonomy also describes a method for applying the taxonomy, first considering the media requirements for the task with basic group and situational characteristics. They suggest this defines the default media requirements for the task. By then considering how the situation and users impact the task, the media characteristics that are required to accomplish the task can be refined.

2.3.4.6 Media Richness Theory

Media richness theory (MRT) [DL86] and other media selection theories such as Media Synchronicity Theory (MST) [DFV08] are focused on choosing the right communication media for the task being undertaken. Like many CSCW theories and frameworks, these theories assume that face-to-face communication is the *richest* communication medium. Although originally targeted at exploring how communication is carried out within

organizations, MRT and MST can be applied to the study of both collocated and distributed synchronous collaboration. In particular, the concept of the *richness* of the media and how that media can be used to meet the needs of a collaboration task help to bridge the task/technology gap.

MRT suggests a continuum of communication mediums in order of decreasing richness: face-to-face, telephone, personal documents, impersonal documents, and numeric documents. One important contribution of MRT is its consideration of equivocality (ambiguity) and uncertainty as dimensions with which to consider communication. MRT implies that communication that involves ambiguous information can be facilitated by rich media tools that communicate multiple cues such as tone of voice and body language. Similarly, lean media are less appropriate for resolving equivocal issues but are more effective at solving problems that involve uncertainty [DL86]. Although MRT has been challenged in its application to synchronous, distributed collaboration [DFV08], the task characteristics of equivocality and uncertainty remain important in determining appropriate media to accomplish a collaboration task.

Rather than considering equivocality and uncertainty, MST suggests that communication consists of a combination of two processes, conveyance of information and convergence of meaning [DFV08]. Similar to Shannon and Weaver's model of communication [SW49], MST assumes these processes require both transmission and decoding of information. Although many communication theories consider the characteristics of media for the transmission and decoding of information, most of these characteristics are socially derived (immediacy of feedback, social presence). MST attempts to characterize media characteristics at a level that is specific enough for testing as well as capable of denoting a range of impacts on communication performance. MST also suggests that tasks are often composed of a set of serial sub-tasks, that different media may be required for each sub-task, and that using more than one media type may be beneficial for a single sub-task.

2.3.4.7 Theory of Remote Scientific Collaboration

Of particular relevance to the study of distributed, artifact-centric, scientific collaboration are those theories and frameworks that consider scientific collaboration specifically. Given that scientific collaboratories are relatively new phenomena, the study

of scientific collaboratories is also relatively recent (see Section 2.1.1). An important outcome from the Science of Collaboratories project is the Theory of Remote Scientific Collaboration (TORSC), which considers scientific collaboratories from an organizational perspective [OHB+08]. TORSC defines a number of levels of success for scientific collaboratories, including effects on the science itself, effects on science careers of the participants, effects on learning and science education, effects on inspiring others, effects on funding and public perception, and effects on the levels of new tool use. TORSC also defines a set of factors that lead to success, including defining the nature of the work, establishing common ground across the collaboratory, the management process (including planning and decision making), and technology readiness. In particular, the authors suggest that *"…there are opportunities to improve collaboration support by exploring technologies that create tools targeted to specific social processes as a way to supplement the shortcomings of using general-purpose tools alone…"* [OHB+08]. Although TORSC does not explicitly provide tools with which to investigate task-specific processes, the use of TORSC with the frameworks presented above does provide a useful grounding for research into distributed, scientific collaboration.

2.3.4.8 Frameworks for Interactive Visualization

Visualization, the process of generating understanding from images of data, is a critical process in the computational sciences [JMM+06], and therefore it is important to consider visualization frameworks in our research. It is worth noting that the creation of the visualizations AND the interactions with the data to create those visualizations are both fundamental to the visualization process [IHH+10]. It is this exploratory process that leads to insight and understanding. Such interactions are of particular importance when the exploration takes place in collaboration with other researchers. The research presented in this dissertation is focused on exploring the impact that distance has on the collaborative exploration of complex scientific data.

A number of frameworks for visual information analysis have been put forward. Shneiderman identifies a taxonomy of information visualization based on the type of data (dimensionality, temporality, and structure) and the type of tasks performed on that data [Shn96]. Shneiderman suggests seven abstract, high-level tasks for information exploration: overview, zoom, filter, details-on-demand, relate, history, and extract.

Shneiderman points out that it is important to expand and refine these tasks as an important next step in the exploration of this domain. Despite this recommendation, Shneiderman's basic taxonomy is widely cited in information visualization [CC05]. Craft and Cairns critique this high-level use of the Shneiderman's taxonomy, suggesting that because the process decomposition is at a high-level *"… these same characteristics are problematic for theorists who are involved in understanding and precisely describing models"* [CC05]. Craft and Cairns also point out that as of 2005 the taxonomy had not been extensively studied and suggest that the guidelines need to be validated in more detail [CC05]. Tory and Moller present a different taxonomy to that presented by Shneiderman, suggesting a taxonomy that divides visualization techniques based on whether the object of study represents a discrete or continuous entity and how the visualization designer chooses the display attributes [TM04]. Tory *et al.* discuss a set of visualization tasks that focus on what the user is searching for, including spatial relationships, numeric trends, patterns, item details (filtering), and connectivity relationships.

Other researchers have explored the temporal relationship of the tasks performed during interactive visualization. Interactive visualization is a process that occurs over time, and it is therefore important to understand this process if we are to design effective tools for visualization. Early work in scientific visualization by Upson *et al.* describe visualization as a three stage process, including filtering data into subsets, mapping data into visual elements, and rendering data to present a view of that data to the user [UFK89]. Shneiderman's task list (described above) extends this to include seven visualization processes [Shn96]. Mark *et al.* describe a five stage collaborative visualization process where collaborators parse questions, map a variable to a representation, choose a visualization, validate the visualization, and validate the entire answer [MCK03]. The three middle stages (mapping variables, choosing a visualization, and validating the visualization) are described as an iterative process which eventually converges to validation of an answer. The authors use this model to describe the interactions in a user study as well as to uncover process differences between the collaboration tasks considered in the study (a focused question task versus an open discovery task). Park *et al.* describe a similar set of activities during their study of

distributed collaborative visualization in an immersive virtual reality environment [PKL00]. The authors describe the common activity pattern that they observed during their study. These activities include problem interpretation, agreement on visualization tools to use, independent search and adjustment of parameters, reporting of discoveries, and negotiating a conclusion based on findings.

Other important work in this area is Card *et al*'s Knowledge Crystallization process [CMS99, Car08, p. 539]. Card *et al.* describe the visualization process consisting of four main activities, each with a set of operations. These are:

- Acquire information
  - Monitor, search, capture
- Make sense of it
  - Extract information, fuse information, find schema, recode into schema
- Create something new
  - Organize information, create new constructs
- Act on it
  - Distribute results, describe results, perform an action based on results

This approach is noteworthy because it is one of the few frameworks in information visualization that refines the high-level tasks into a set of lower level operations that may be carried out to accomplish those tasks.

Isenberg *et al.* take a slightly different approach to studying the information visualization process [ITC08, NTC07]. Like Tang's early work in studying the group design process in a physical tabletop environment [TL88, Tan89] (see Section 2.3.1.1 for more details), Isenberg *et al.* study the process of collaborative visualization by studying how individuals interact with physical artifacts (object-centric collaboration) in a non-digital environment (on a physical tabletop). Their study explores interactions unconstrained by technology, with the goal of getting a better understanding of how individuals understand and think about the problem. Like the frameworks discussed above, the authors identify a set of processes that study participants carried out during their study. These include browsing through data, parsing task descriptions, discussing collaboration approach, establishing a task strategy, clarifying information about an artifact, selecting artifacts pertinent to a task, operating (extracting information) on (from) artifacts, and validating or confirming knowledge or a solution to the task.

Of particular importance from this study are the author's findings that although some tasks typically occur before other tasks, there are no overall temporal patterns among the tasks. In addition, the authors found that individuals varied in how they approached the tasks. These findings are important in that they reinforce previous findings in terms of the types of tasks that are performed in information visualization. Explicit comparisons between the author's tasks and those presented by Card *et al.* [CMS99, Car08], Mark *et al.* [MCK03], and Park *et al.* [PKL00] are given in [NTC07]. These results are also important because they suggest that consistent temporal patterns of specific collaboration tasks may be relatively rare across tasks and individuals [ITC08].

2.4 Summary

This chapter presents an overview of the foundational research areas that are relevant to the domain of distributed, artifact-centric, scientific collaboration. Initially, we discussed current research into collaboration in the sciences, considering the domain of computational science, scientific collaboratories, and data-centric science. We then explored the science of collaboration, considering a broad range of related research areas, including communication, social psychology, gesture, and cognitive psychology. Related research in the domain of Computer Supported Collaborative Work (CSCW) is then discussed in detail. Lastly, we considered how all of these domains are inter-related by exploring theories, models, and frameworks that tie this research together. The exploration of these research areas provide us with the broad base of knowledge that is required to perform the research carried out in the dissertation, while at the same time pointing out gaps in these research domains where the research presented here contributes new knowledge.

Part II - Methodology

3 Research Approach

The study of group work, and in particular distributed scientific collaboration, is a complex research domain. In Chapter 2, we discuss the widely varied research areas (sociology, psychology, communication, linguistics, computer science) that impact this area. It is not surprising that the research methodologies used across these areas vary widely. In this chapter, we provide a high-level overview of research methods in general, with a focus on the methods utilized in this research. We then outline the specific research methodology used in our study of distributed, scientific collaboration. Since the focus of this research is on advanced collaboration technologies, we also briefly discuss the technology assumptions that are prevalent throughout this dissertation.

3.1 Research Methods

One of the key questions faced by a researcher studying human subjects is the type of research methods one chooses to utilize for that study. In the past, there has often been a chasm between the domains of quantitative and qualitative research, with the two being cast as "good research" and "bad research" by the "opposing side." For an interesting and entertaining discussion of the supposed quantitative versus qualitative dichotomy, see Bavelas' paper "Quantitative versus Qualitative?" [Bav95]. Although there is a fundamental difference between the two approaches, it is important to recognize both as valuable methods that can contribute to our knowledge. Indeed, over the past fifty years, mixed method strategies of enquiry that blend both qualitative and quantitative techniques have evolved as an important research method [Cre03, p. 15]. We consider each of these strategies of enquiry below.

3.1.1 Quantitative (empirical) Methods

The approach used in this research is strongly grounded in the scientific method. The Merriam Webster dictionary [Merriam] defines the scientific method as the *"principles and procedures for the systematic pursuit of knowledge involving the recognition and formulation of a problem, the collection of data through observation and experiment, and the formulation and testing of hypotheses"*. This is the domain of the formal (math, logic) and natural (physics, chemistry, etc.) sciences. Researchers in these areas use theory to develop hypotheses, perform empirical observations to test those hypotheses, and use the

results of those observations to support or refute the hypotheses and therefore the theory. The scientific method uses deductive reasoning and quantitative research methods such as experiments (random assignment of subjects to conditions), quasi-experiments (experiment with some assumptions, often random assignment, not met), and correlational studies [Cre03, p. 13]. The scientific method reflects a post-positivist approach to knowledge. Creswell states that post-positivism and in turn the scientific method make the following key assumptions [Cre03, p. 7]:

- That knowledge is conjectural and that the absolute truth can never be found;
- That research is the process of making claims and then refining or abandoning those claims for others that are more warranted;
- That data, evidence, and rational considerations shape knowledge;
- That research seeks to develop relevant true statements, ones that can serve to explain the situation that is of concern or that describes the causal relationships of interest; and
- That objectivity is essential to post-positivist enquiry.

3.1.2 Qualitative (exploratory) Methods

Although strongly rooted in a post-positivist scientific method, our study of distributed scientific collaboration is also strongly influenced by the social sciences. The social sciences (psychology, sociology, etc.) also make use of empirical science. At the same time, they often take a more social constructivist approach to knowledge, recognizing that individuals develop subjective meaning from their experiences and that the basic generation of meaning is social and arises from interactions with the world [Cre03, p. 9]. As a social constructivist researcher, the goal is to interpret the meanings others have constructed about the environment around them. Such an approach typically relies on inductive reasoning (rather than the deductive reasoning used in the post-positivist based scientific method) to develop theory from observations. When using inductive reasoning, one begins with specific observations and measures, detects patterns and regularities from those observations, formulates tentative hypotheses that can be explored, and develops general conclusions and theories. Creswell, reflecting earlier research from others, suggests constructivism is based on the following key assumptions [Cre03, p. 9]:

- That meanings are constructed by human beings as they engage the world they are interpreting;
- That humans engage with their world and make sense of it based on their historical and social perspectives; and
- That the basic generation of meaning is always social, arising in and out of interaction with a human community.

Qualitative and constructionist research approaches are commonly used in the social sciences, using techniques such as ethnography, grounded theory, and case studies (we discuss some of these approaches in more detail below). It is important to note that good qualitative research does not simply explore and describe. As stated by Gherardi and Turner, *"What is wanted is not a social 'shopping list' which records what has been noticed, but an account of a series of interactions with the social world in a form which plausibly alerts us to the possibility of a new order not previously seen – a theoretical account"* [GT02].

3.1.3 Mixed (integrated) Methods

Over the last 50 years, many researchers have come to realize that approaching a research area by using a single research methodology is needlessly restrictive. As a formal form of enquiry, mixed method research designs (where both qualitative and quantitative data are gathered) are relatively new. Creswell's previous book on research methods, published as recently as 1994 [Cre94], did not consider mixed methods research, focusing on qualitative and quantitative methods only. In the second edition of Creswell's book, published in 2003, mixed methods of enquiry are considered at the same level as qualitative and quantitative enquiry [Cre03]. Mixed method research is tightly coupled with the pragmatic view of knowledge, where *"Instead of methods being important, the problem is most important, and researchers use all approaches to understand the problem"* and that researchers should *"… focus attention on the problem in social science research and then [use] pluralistic approaches to derive knowledge about the problem"* [Cre03, p. 11]. Creswell summarizes some of the key principles of pragmatic claims about knowledge as [Cre03, p.12]:

- Pragmatism is not committed to any one system of philosophy and reality;

- Individual researchers have freedom of choice to choose the methods and techniques and procedures that meet their needs;
- Pragmatists do no see the world as an absolute unity;
- Truth is what works at the time; it is not based in a strict dualism between the mind and a reality completely independent of the mind;
- Pragmatist researchers look to the "what" and "how" to research based on its intended consequences;
- Pragmatists agree that research always occurs in social, historical, political, and other contexts; and
- Pragmatists believe that we need to stop asking questions about reality and the laws of nature.

The key benefit of the mixed methods approach, and the associated pragmatic view of knowledge, is that it recognizes that all methods have limitations and that using multiple methods helps the researcher to neutralize those limitations. The ability to triangulate with qualitative and quantitative methods allows researchers to avoid bias, inform research design across methods, and provide different levels of analysis. It is this approach that is most consistent with the research presented in this dissertation.

It is important to point out that mixed methods research can be considered at a number of levels. At the study design level, as presented by Creswell [Cre03], one can design a study that uses quantitative, qualitative, and mixed methods strategies of enquiry. Within a study, a mixed methods design may use a sequential (a quantitative method followed by a qualitative method or vice-versa) or concurrent (converge quantitative and qualitative methods to provide a comprehensive analysis of a problem). At a higher level, mixed methods allow individual researchers and the overall research community to triangulate their research, approaching a problem from many directions and providing a much richer and complete view of a problem domain.

3.1.4 Research Methods Summary

The triangulation of research is widely recognized as fundamentally important in the social psychology, gesture, and group works communities. Joseph McGrath suggests that the goal of research in the behavioural sciences is to maximize the generalizability of the

evidence over a population of actors, the precision of measurement of the behaviours, and the realism of the situation or the context [McG84, McG93b]. He also points out that it is not possible to maximize all three with a single research method, stating that: *"a) that methods enable but also limit evidence; b) that all methods are flawed, but all are valuable; c) that the different flaws of various methods can be offset by (simultaneous or successive) use of multiple methods; and d) that such multiple methods should be chosen to have patterned diversity, so that strengths of some offset weaknesses of others"*.

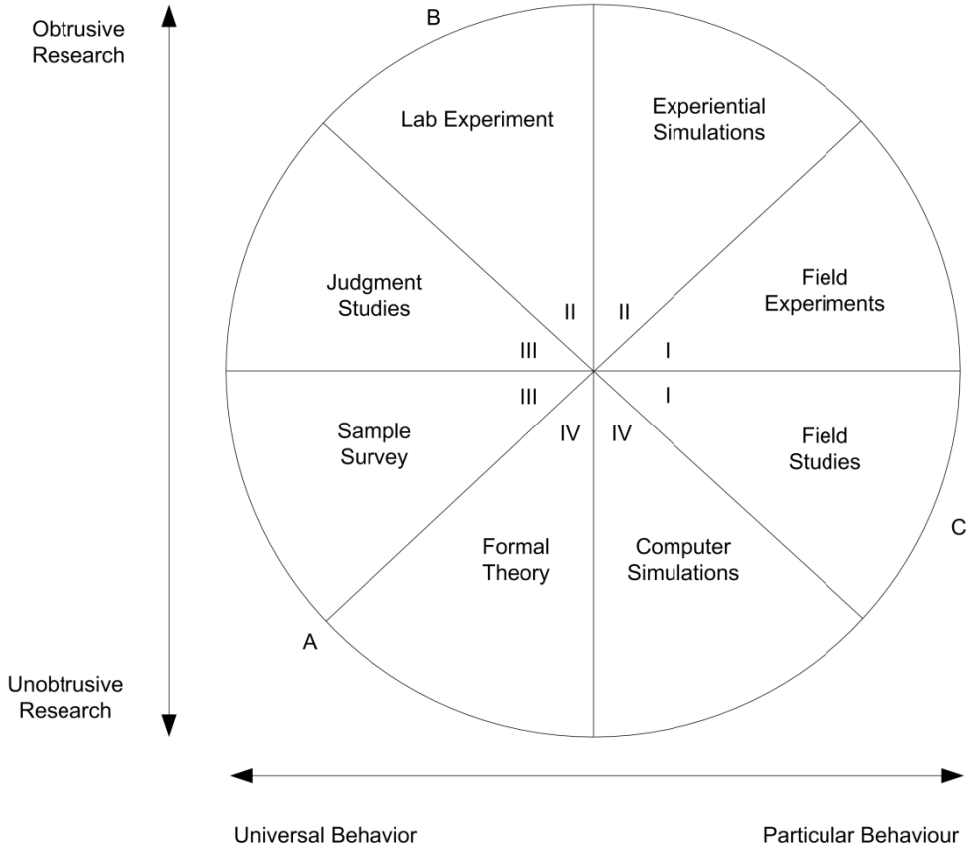


Figure 5: McGrath's research strategies (reproduced from [McG84])

The research choices that we make in choosing a method are diagrammed in Figure 5 (from [McG84]). The vertical dimension represents the obtrusiveness of the research method, while the horizontal dimension represents the level of universality (versus particular use) of the method. The point where there is maximum control over generalizability, precision in control and measurement, and realism of context are represented in the figure by the letters A, B, and C respectively. Methods can also be classified into those that are used in natural settings, used in contrived settings, are setting

independent, or have no observational requirement at all (represented by the octants labelled I, II, III, and IV respectively in Figure 5). McGrath's representation of the spatial relationships in this diagram highlights the dilemma of the social science researcher in picking a research method. For example, the more a research method maximizes the control over precision of measurement (performing a laboratory study, thereby increasing B) the less likely it will be that the method can be applied in a natural setting (decreasing C). Similarly, maximizing the naturalness of the setting (studying a research group in its natural environment, thereby increasing C) will reduce generalizability to the population (decreasing A) or the precision and control over the data that is gathered (decreasing B). Clearly, an approach that triangulates a given research domain, using complementary research methods (qualitative and quantitative approaches), is beneficial in providing a complete understanding of a specific research question.

This approach is widely supported throughout the literature that is relevant to distributed, scientific collaboration. For example, Clark criticizes researchers that study the use of language as being isolated in their methods, with cognitive psychologists studying individuals and social scientists studying group processes. Clark suggests that the integration of these two research domains is essential to acquiring a complete picture of language use in communication [Cla96, Section 2.2.3]. Olson and Olson also stress the importance of using multiple methods in studying group work, stating that *"… we feel it is critical to study such phenomena through a linked approach using both field and laboratory work"* [OO01]. Last, but certainly not least, Janet Bavelas, a prominent researcher in the field of gesture, criticizes false dichotomies in general, and the false dichotomy of quantitative versus qualitative research in the study of gesture in particular [Bav95].

3.2 Research Methodology

The research presented in this dissertation uses a mixed methods approach, utilizing both quantitative and qualitative research methods where appropriate. The importance of using mixed methods to study scientific collaboration has been pointed out by Sonnewald *et al. "Collaboratory evaluation can have multiple purposes and goals. Examples include: increasing our understanding of individual behaviour in geographically distributed collaboration, discovering new knowledge about collaborative scientific work*

processes as mediated by technology, informing the design of collaboratory technology, and providing insights regarding the efficacy of scientific collaboratories. These purposes are complex and multi-faceted, often requiring multiple comprehensive studies that employ qualitative and quantitative research methods" [SWM08].

Our use of qualitative techniques is driven by two key factors. First, the domain of distributed, scientific collaboration is not a well studied field. There are few detailed studies of collaboratory use, with most of the research that has been performed considering broad analyses of collaboratories rather than detailed analyses of how researchers utilize technology for collaboration (see Section 2.1). Secondly, the use of advanced collaboration technologies is rarely studied in a naturalistic environment (see Section 2.3). This is true in both the general case, as well as in the case of distributed, scientific collaboration. It therefore was necessary to perform a set of exploratory studies (Chapter 5 through Chapter 7) to ground our later quantitative research (Chapter 8 through Chapter 11).

We make use of quantitative techniques in a number of ways. We use a range of quantitative measures throughout our qualitative studies. Thus, all of our studies have a quantitative dimension to them. More importantly, our exploratory studies (Chapter 5 through Chapter 7) provide a broad basis of understanding of how scientists use advanced collaboration technologies. Worthy of note is the fact that the studies presented in Chapter 6 and Chapter 7 are longitudinal studies of scientific researchers in both naturalistic and advanced technology environments. Such studies are extremely rare, providing us with a unique view of how scientists collaborate. This provides us with the foundation for the design of a quantitative laboratory study of the use of human communication channels in distributed, scientific collaboration (Chapter 8 and Chapter 11).

The research presented in this dissertation uses three basic research strategies: case studies, ethnographies, and experiments. We define and briefly discuss our use of each of these research methods below.

3.2.1 Case Studies

Creswell defines a case study as a research strategy *"… in which the researcher explores in depth a program, an event, an activity, a process, or one or more individuals.*

The case(s) are bounded by time and activity, and researchers collect detailed information using a variety of data collection procedures over a sustained period of time" [Cre03, p. 15]. The goal of a case study is not to find cause and effect, but instead to explore, describe, generate new propositions, and build new theory. Case studies were used multiple times throughout this research. We use case studies to explore the application of existing theory in new contexts (scientific collaboration) as well as to reveal new insights and generate new hypotheses about how advanced collaboration technologies are used. These case studies help us to reach the following research objectives:

- *Objective 1: Develop a broad understanding of how scientific researchers collaborate.*
- *Objective 3: Evaluate advanced collaboration modalities and technologies for scientific collaboration.*

The first case study explores the use of CoTable, a collocated and distributed collaboration technology prototype (Chapter 5). The goal of this research was exploratory in nature, providing us with new insights about how advanced collaboration technologies are used. Many of the hypotheses explored in Chapter 8 through Chapter 11 originated from this case study. We explore the use of the CoTable system with a small number of users, employing participant observation and interviews to gather data about user experience.

We also performed a case study of the use of Scientific Media Spaces (SMS) in the support of distributed, scientific research at the IRMACS Centre (Chapter 6). This case study consisted of a detailed analysis of the use of the SMS infrastructure at IRMACS over a five year period (2005 – 2009). Usage statistics and participant surveys were used to gather data about the use of the collaboration technology.

Smaller scale case studies were also performed as part of other analyses, including individual case studies of collaboration scenarios that complemented other research. In particular, we utilized the gesture coding scheme developed and used in the ethnography performed in Chapter 7 to perform a case study analysis of a distributed research seminar and the video presentation used in our laboratory study of gesture from Chapter 8 and Chapter 9.

3.2.2 Ethnographies

Creswell defines ethnography as a research strategy where *"… the researcher studies an intact cultural group in a natural setting over a prolonged period of time by collecting, primarily, observational data. The research process is flexible and typically evolves contextually in response to the lived realities encountered in the field setting."* [Cre03, p. 14]. Ethnography is an important research tool in the anthropology and sociology fields. The goal of ethnography is to study a community of practice in a natural setting, providing an insider's view of, in our case, how scientific researchers collaborate.

In Chapter 7, we perform a longitudinal (five month) ethnography of a group of scientific researchers as they collaborate in a natural work environment. This ethnography is critical to this research, as it provides us with valuable data on how researchers collaborate using digital data. Equally importantly, it also provides us with data on how those same researchers use advanced technologies such as touch-sensitive interaction devices (Smartboards) and distributed collaboration technologies. The study helps to meet the following research objectives:

- *Objective 2: Develop a deep understanding of how scientific researchers interact with digital artifacts when they collaborate.*
- *Objective 3: Evaluate advanced collaboration modalities and technologies for scientific collaboration.*

This study addresses both of the issues presented by Sonnewald [SWM08] and Heath and Luff [HL91] mentioned in Section 3.2. Sonnewald points out that the naturalistic study of collaborating scientists is important in the study of collaboratories. Heath and Luff point out that it is difficult to draw technology design implications from the naturalistic studies that have occurred in the CSCW literature because the studies involve technologies that are very different from those being developed in the CSCW community. This study allows us to study both collaborating scientists AND how they use advanced technologies in a naturalistic environment.

3.2.3 Laboratory Experiments

Creswell defines experiments as including *"… true experiments, with the random assignment of subjects to treatment conditions, as well as quasi-experiments that use non-randomized designs"* [Cre03, p. 14]. Experiments are carried out under controlled

conditions and are targeted at testing hypotheses and discovering causality. They are at the other end of the qualitative/quantitative continuum from our exploratory ethnography and case studies discussed above. At the same time, they complement this research and help us to meet the following research objectives:

- *Objective 2: Develop a deep understanding of how scientific researchers interact with digital artifacts when they* collaborate.
- *Objective 4: Evaluate advanced collaboration modalities and technologies for scientific collaboration.*

Building on the data gathered from our case studies and ethnography, we designed a laboratory experiment that answered some key questions about distributed, artifact-centric, scientific collaboration. In particular, our qualitative explorations suggest that researchers' collaboration practices change in the presence of technology, that digital artifacts are an important part of scientific collaboration, and that gesture is used extensively in referring to those digital artifacts. One key question that is not answered by these studies is whether or not collaborating scientists attend to those gestures as part of the decoding and understanding process of communication. The study presented in Chapter 8 and Chapter 9 provides new knowledge about how researchers decode and understand information that is communicated about digital artifacts.

3.3 Multi-dimensional research approach

The study of distributed, scientific collaboration is a multi-faceted and complex research domain. Our approach to studying this area takes a multi-dimensional approach:

- Quantitative/Qualitative: As discussed above, we utilize both quantitative and qualitative research methods.
- Macro/Micro: We analyze the use of advanced collaboration tools at the macro-level (use by a large research community over a five year period) and the micro-level (use by a small research group over a five-month period).
- Co-located/Distributed: We analyze the use of both collocated (collaborators in the same room) and distributed (collaborators at two or more distributed locations) collaboration.

- Encoding/Decoding: We analyze both the encoding (how information is sent) and decoding (how information is received) processes researchers use to communicate about complex scientific topics.
- Prototype/Production: We analyze the use of state-of-the-art technical infrastructure in both research prototype (experimenting with new HCI and CSCW technologies) and production (observing active researchers using sophisticated CSCW tools) environments.

3.4 Technology assumptions

One of the design goals for the distributed, object-centric systems presented in the papers by Kraut *et al.* [OFC+03][KGF02][FKS00][KMS96] (as discussed in Section 2.3.3.1) was to provide a simple cost effective system using commodity technologies. This is similar to the approach suggested by Whittaker and O'Conaill in that they suggest it is necessary to understand the use of low-quality video in the design of collaboration tools because of limited bandwidth at the time [WC97]. Although this is necessary when considering deployment of tools on contemporary technologies, we take a different approach. Our goal is to communicate the information that is required to perform a specific task as effectively as possible. We make use of advanced interaction, display, and networking technologies where appropriate to maximize the "quality of experience". We chose this approach intentionally. If commodity technology is being used, we are restricted to asking the question "Given that we have technology A and B, how can we provide the best collaboration for task T?" We do not want to be limited by the technology, but instead want to be able to ask the question "How can we provide the most effective collaboration for task T?" These are fundamentally different questions.

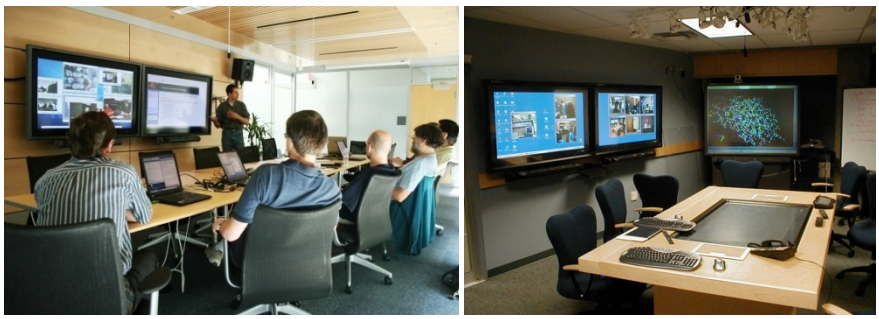


Figure 6: Example advanced collaboration environments

We assume that the technologies we utilize will be available to our users in the near future at a reasonable cost. This is particularly likely in the scientific research environment, where collaborators have access to advanced networking, sophisticated visualization environments, and advanced interaction technologies as part of the research infrastructure that is available to them at their academic institutions.

The scientific research environment that we are targeting our collaborative tools at is a sophisticated one. A brief example illustrates this. WestGrid (www.westgrid.ca) is a large-scale grid-computing consortium that spans four provinces and fourteen research institutions in Western Canada. In addition to the high performance computing infrastructure that is common for such a project, WestGrid institutions have also built an extensive collaboration and visualization infrastructure. Each institution has created an advanced collaboration room that allows scientific researchers to collaborate with distant colleagues.

Each room consists of two to four displays (projectors, plasmas, tabletop displays, etc.), two or more cameras, high quality audio, and in some cases advanced interaction (touch screens) and visualization technologies (stereoscopic displays). Two examples of such advanced collaboration rooms are shown in Figure 6. These rooms typically support a wide range of collaboration technologies, including traditional teleconferencing, video conferencing (often high definition), and desktop collaboration technologies such as AccessGrid [COP+00], iChat [Car03], Skype [MR06], and VNC [RSW+98]. Many such rooms make use of touch sensitive screen overlays (Smartboards [Smart]), allowing users to interact with applications by directly touching the screen or annotating documents by writing on the screen with a digital pen. Institutions are typically connected together by a dedicated gigabit network that enables high throughput, low latency data transfers. This is the technology environment in which we perform our studies, the environment in which our technology prototypes are built, and the environment for which our design guidelines are targeted. Given that such a technology environment is available today to the academic research community, we assume that these technologies will be commonly available to the wider community in the near future.

4 CoGScience – A New Collaboration Framework

In this chapter, we present a conceptual framework for studying distributed, artifact-centric, scientific collaboration. We call this framework the Collaborative Group Science (CoGScience) Framework. One of the unique aspects of this framework, as described below, is its integration of both social science (communication, social psychology, and cognitive psychology) and computing science (HCI and CSCW) into the framework. In particular, it bridges the gap that exists in existing frameworks between the human communication needs of artifact-centric scientific collaboration and the technological aspects of how those communication needs can be delivered. The CoGScience name incorporates these fundamentals through a play on words, using CoG to imply the social (**CO**llaborative **G**roup work), cognitive (**COG**nitive science), and technical (**COG** as a part of a mechanical system) aspects of artifact-centric scientific collaboration. Similarly, the word **Science** refers to both the application domain to which the framework is applied (collaborative computational science) as well as the scientific basis on which it is built (communication, social psychology, cognitive psychology, and computer science).

The development of this framework was motivated by the issues that arose during the design, implementation, and assessment of the CoTable collaboration system (Chapter 5). The need for a new framework was reinforced as our artifact-centric collaboration ethnographic study (Chapter 7) was designed, performed, and analyzed. None of the existing frameworks were able to capture the depth, breadth and subtleties that we were discovering during the CoTable and ethnographic studies. We addressed the shortcomings we found in existing frameworks with the creation of the CoGScience Framework.

It is impossible to discuss both the CoGScience Framework and our CoTable and ethnographic study in chronological order, as they occurred in parallel. Indeed, the framework informed the design of the studies and the design of the studies advanced the development of the framework. The CoGScience Framework is presented first, as we use it as a tool in later chapters. We refer extensively to the CoTable system and the ethnographic study throughout this chapter.

4.1 CoTable Overview

The goal of creating the CoTable system was to provide experience with emerging technologies and how they might be used to support collocated and distributed collaboration. CoTable was created not to improve on these systems, but to provide a test bed in which we could gain experience with both advanced technologies and advanced interaction techniques. This included exploring touch sensitive tabletop display and interaction devices, gesture-based user interaction, multiple displays, multiple cameras, and remote audio and video collaboration technologies.

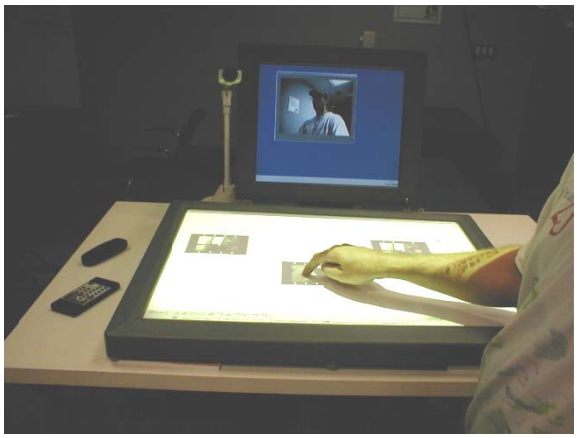


Figure 7: The CoTable system in action

The goal of creating the CoTable system was to create a collaboration environment that attempts to replicate the rich face-to-face collaboration environment that we experience when working together using digital artifacts on a tabletop. When creating the distributed components of the system, we wanted to maintain as many communication channels as possible. From Buxton's perspective, we attempt to provide a personal space, a task space, and a reference space [Bux09]. Recall that Buxton's defines personal space as the space where one sees personal communication cues such as facial expression, gaze, and emphatic gestures while task space is where one sees and interacts with the work object. Buxton's reference space fuses the personal and task spaces, resulting in a space where remote collaborators can directly see personal interaction with the work object. Most remote collaboration environments provide one or two of these spaces, typically as distinct communication streams. Rarely do systems provide all three spaces, and even when they do one or more of them are typically highly impoverished in the amount of information that they communicate (see Section 2.3.3.2 for details).

The CoTable system implements a personal space by providing a view of the remote user's face across the tabletop (a personal visual stream), a shared task space through a multi-user collaborative tabletop application (a task visual stream), as well as a shared reference space through the use of a video feed of the tabletop that shows intentional action of gestures combined with the tabletop (a reference visual stream – not shown in the image above). An aural stream is also provided for verbalization. For an in-depth description of CoTable, refer to Chapter 5. An image of the CoTable system in action is shown Figure 7.

4.2 Genesis of the CoGScience Framework

In order to distil research into a coherent body of work that can be used to design a new collaboration system, it is necessary to find a common ground on which to consider the issues. Group work frameworks provide such a common ground (see Section 2.3.4 for a detailed discussion of existing frameworks). The realization that current group work frameworks did not address all of the needs of this research came about during the design, implementation, and testing of the CoTable system.

A review of collocated and distributed collaboration research reveals that little research had been carried out on the impacts of distance on collaborators working in a technology mediated tabletop environment. The goal of creating the CoTable system was to provide experience with emerging technologies and how they might be used to support collocated and distributed artifact-centric collaboration. In attempting to understand this complex design space, we utilized a number of existing frameworks in the design, implementation, and analysis of the CoTable system. In particular, we utilized the Mechanics of Collaboration [GG00, Section 2.3.4.4] and the ETNA Taxonomy [MJA+02, Section 2.3.4.5].

As the CoTable design proceeded, it was realized that although these frameworks were very useful, they did not address a number of our key questions. The gaps discovered in these frameworks demonstrated the need for the development of the CoGScience Framework. As research proceeded and we considered the design and analysis of our artifact-centric collaboration ethnographic study (Chapter 7), we also began to make use of the CREW Framework [0097, Section 2.3.4.3]. Although the CREW Framework provided another perspective on this problem domain, it also failed to capture all of the

subtleties that we encountered during the CoTable and ethnographic studies. This further motivated the development of the CoGScience Framework.

In the following sections, we consider the benefits and the limitations of the frameworks we considered and point out the issues we discovered pertaining to this research. We then present the CoGScience Framework, a new framework that addresses some of these limitations.

4.3 Applying Existing Frameworks to Tabletop Collaboration

The discussion presented in this section focuses on our use of frameworks during the development of the CoTable system. We used a four-step process in designing, implementing, and assessing the CoTable system. We first tabulated the needs of collocated artifact-centric collaboration on a tabletop interaction device. This analysis was primarily done using MOC. We then attempted to identify the information that would be lost when the collaboration occurred at a distance. For this step we used our needs analysis from MOC in tandem with the technology aspects of ETNA. We then implemented mechanisms that provide some (or all) of the information that is lost to the distributed users. Finally, we analysed the implementation. We used a combination of a MOC needs analysis and the ETNA technology characteristics to attempt to determine the effectiveness of the mechanisms that we used to communicate information to remote collaborators on the CoTable system. Note that we do not present this analysis in this chapter, but instead focus on the gaps we discovered in the MOC, ETNA, and CREW frameworks. A detailed analysis of CoTable, using the CoGScience Framework, is given in Chapter 5.

4.3.1 The Mechanics of Collaboration and CoTable

We initially applied the Mechanics of Collaboration (MOC) [GG00, Section 2.3.4.4] to our design of the CoTable system in order to determine the communication needs for artifact-centric collaboration on a tabletop device. Note that this process can be done independent of the CoTable system. We are concerned with processes or activities that need to be performed to accomplish the task. Gutwin defines seven major activities for MOC: explicit communication, consequential communication, coordination of action, planning, monitoring, assistance, and protection.

Explicit communication occurs when group members intentionally provide others with information through verbal, written, or gestural communication. *Verbal communication* is an important component of a collaboration environment, and therefore an explicit *verbal channel* between the remote collaborators was deemed necessary. Since artifact-centric collaboration is a visual task where users refer to artifacts, deictic references ("this one") combined with pointing to an artifact can be an important part of the communication. Thus an explicit *gestural channel* was also deemed necessary.

Consequential communication occurs when users communicate information unintentionally. This type of interaction is very prominent in an artifact-centric collaboration because of the need to communicate information about the artifact. This subtle aspect of communication can be very important, and we need to consider communication that includes *facial expression, body language, mannerisms*, and *consequential gestures*. Thus it was deemed necessary to provide both *facial* and *gestural consequential communication channels*.

Coordination of action, planning, monitoring, assistance, and protection are all highly important activities in artifact-centric collaboration. Failure to provide mechanisms that communicate such activities can result in conflict, awkwardness, and duplicated action. A *workspace awareness channel* that provides the ability to communicate these activities was also deemed necessary.

Using MOC provides us with a list of activities that should be considered for artifact-centric tabletop collaboration. We have mapped those activities to a set of channels that need to be communicated to remote participants. Note that the concept of a channel that communicates information for a purpose does not exist in MOC, and the channels defined above are mechanisms we use to capture the ways in which the MOC activities can be communicated to remote collaborators. In fact, MOC does not consider task, channel, or technology, limiting its usefulness beyond the analysis presented above. Note that Gutwin and Greenberg do not claim that MOC does more than this, stating that MOC is necessary, but not sufficient, to understand complex collaboration tasks.

4.3.2 The ETNA Taxonomy and CoTable

ETNA considers the impact that task, user, and environmental characteristics have on the media that are used to communicate information to remote collaborators [MJA+02,

Section 2.3.4.5]. Like MOC, ETNA is useful in helping us scope our research. Like MOC, it is also incomplete for our purposes.

The fundamental division that ETNA utilizes is whether a task is person (telepresence) or data (teledata) focused. Although ETNA considers a rich set of task, user, and environmental characteristics for telepresence, the characteristics that are considered for teledata are relatively impoverished. This is severely limiting when considering artifact-centric collaboration scenarios. Using Buxton's terminology, artifact-centric collaboration requires a mix of personal (telepresence) and task space (teledata), with the intersection of the two, Buxton's reference space, being fundamental to the process [Bux09]. That is, most artifact-centric collaboration systems, such as CoTable, are hybrid telepresence and teledata systems. ETNA, with its limited teledata characterization, is therefore restricted in how it can be used to study artifact-centric collaboration.

Although the ETNA taxonomy has a media focus and is one of the few frameworks to apply media characteristics to task in a rigorous fashion, it does not consider media characteristics consistently across the task, user, and environmental characteristics. Similarly to the impoverished task, user, and environment characteristics, ETNA again limits the characteristics it considers for media in teledata environments. In essence, the ETNA taxonomy suggests that we know enough about personal and task space to only consider specific technologies for these high-level task dimensions. We believe that the ETNA taxonomy does not consider a wide enough range of technology characteristics (for task space or teledata in particular) and that it is premature to encode such restrictions into the framework. As with MOC, such restrictions make it difficult to apply ETNA to an artifact-centric collaboration system.

Despite these limitations, we applied ETNA by considering the telepresence characteristics for task, user, and environment characteristics in our analysis of the teledata media streams used in the CoTable system. This approach is justifiable because many of our teledata streams have a strong telepresence component. That is, in the CoTable task stream (the overhead camera), we can see both personal space and task space (or a reference space). Although this is not using ETNA as intended, it is the only practical way to apply ETNA to an artifact-centric system. Note that we apply the

CoGScience Framework to the CoTable technology characteristics in much the same way (Section 5.4).

Although ETNA has some limitations, it is still a very powerful tool. It is the most complete framework that we are aware of for considering the impacts of the task, user, and environment characteristics on media communication. In addition, MOC and ETNA are complementary. While MOC considers activities but not task or technology, ETNA considers task and technology but not activities. Despite these limitations, combined MOC and ETNA allow us to perform a relatively detailed analysis of the CoTable system.

4.3.3 The CREW Framework and CoTable

Perhaps the most complete framework for studying group work is that proposed by Olson and Olson, which we call the CREW Framework [OO97, OO01, Section 2.3.4.3]. Like the ETNA taxonomy, it is task focussed. Unlike ETNA, it focuses on the cognitive aspects of the task and considers a range of characteristics including the nature of the material (ETNA's teledata/telepresence), the major information processing activity (MOC's planning, helping, etc.), dependency of group members (ETNA's user dimension), the mental resources required to perform the task (ETNA's cognitive component), and the duration and scope of the task (ETNA's situation dimension). The CREW Framework also considers the technological dimension, dividing it up into technologies that support the conversation and technologies that support the work object (ETNA's telepresence and teledata dimension and Buxton's personal/task/reference spaces). The CREW Framework therefore integrates much of the ETNA and MOC frameworks into a single, comprehensive framework.

Despite this apparent integration of the two frameworks, it does not meet all of our needs. Firstly, the CREW Framework as described above (and as described in Section 2.3.4.3) is actually our distillation of the key aspects of two slightly different descriptions of the CREW Framework [OO97][OO01]. Thus the big picture presented above is our view of how these two papers fit together. A single presentation of the CREW Framework, other than as described above, does not exist.

There is one other important gap in the CREW Framework. The CREW Framework does not consider the technology and media characteristics to the same degree as does the

ETNA taxonomy. In particular, the CREW Framework does not consider concrete technologies in detail, but rather treats the requirements of the task as higher level processes such as gaining access to a shared artifact, turn taking, reciprocity of communication, and coordination. Technology characteristics are discussed briefly (video clarity, field of view, audio clarity, delay, and control) but only in passing before going on to consider the higher level processes that the technology supports. When considering the CoTable system, and our analysis of how users use the system, it is a consistent mapping from task (artifact-centric collaboration), to process (coordination of action), to communication channel (gesture), to technology (video stream of the workspace) that is missing from all of these frameworks.

Throughout our analysis of the CoTable system, none of the frameworks we utilized provided a complete solution. Rather than attempt to use multiple frameworks to analyze different aspects of the same collaboration environment, we created a new framework that embodied all of these elements. This framework is called the CoGScience Framework.

4.4 CoGScience: A Framework for Artifact-Centric Collaboration

We utilize the frameworks described above as the basis for the CoGScience Framework. In particular, our starting point is the CREW Framework [OO97, OO01] as it embodies a number of key components from other frameworks. The CoGScience Framework also leverages the fundamental building blocks of the communication models (see Section 2.2.1) and task frameworks (see Section 2.3.4) provided by the Communication and Group Work research communities. In particular, we take a similar stance to that suggested by Dennis *et al.* in Media Synchronicity Theory [DFV08] and Shneiderman [Shn96]. That is, we believe that the granularity with which task and technology are considered in the frameworks and theories discussed previously (Section 2.3.4 and Section 4.3) do not consider these dimension in an appropriate level of detail for the analyses that are required to effectively explore this problem domain. The CoGScience Framework is also influenced by previous work carried out by the author and colleagues on the impacts of task and technology on the quality of experience in advanced collaboration environments [PSC+04][CZP+05].

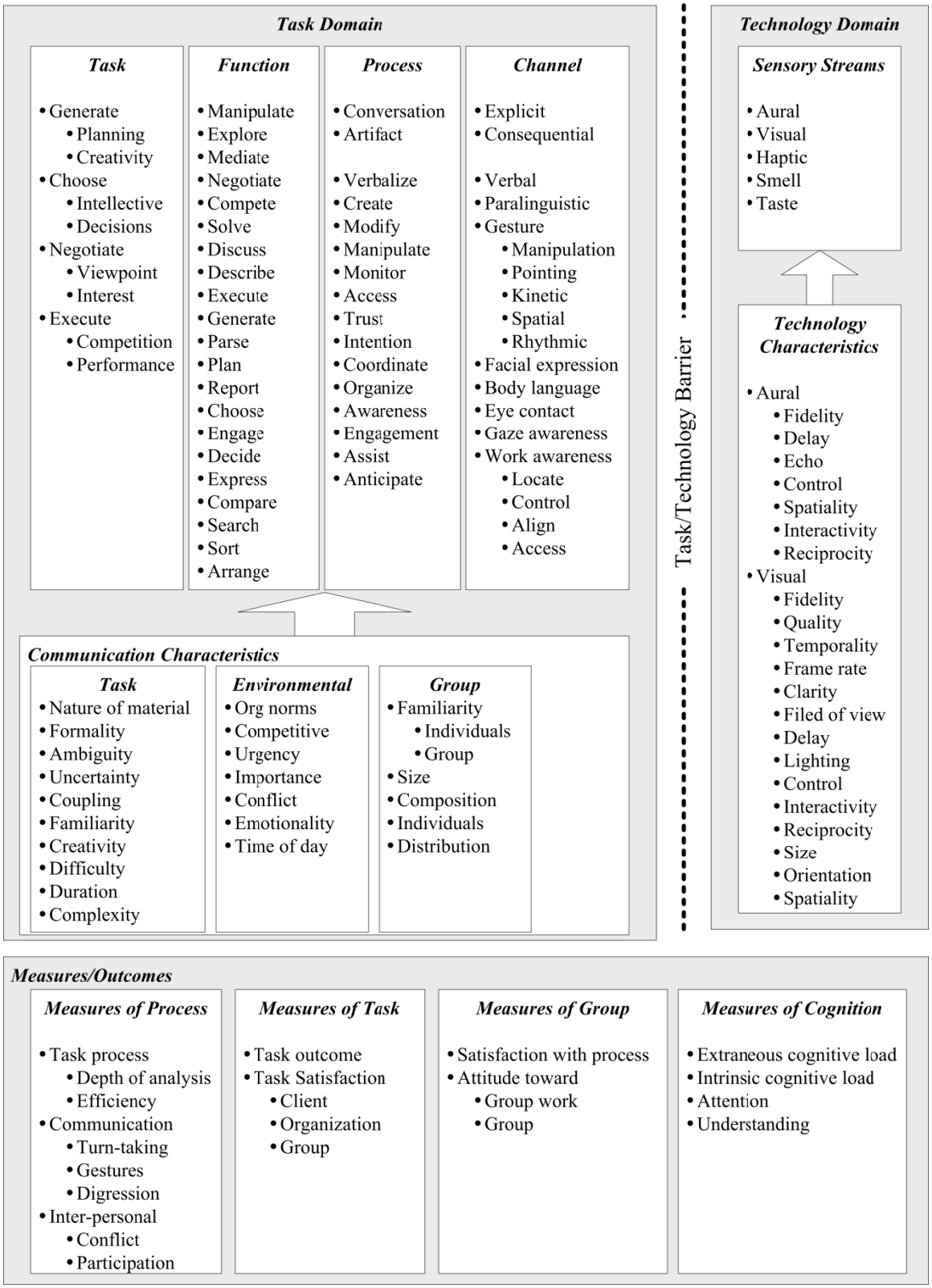


Figure 8: The CoGScience Framework

In considering the richness of the artifact-centric collaboration domain, the CoGScience Framework is necessarily complex. A diagrammatic representation of the framework is presented in Figure 8, and we refer to this figure extensively. In order to assist in understanding the discussion presented in the following sections (and indeed in other chapters), whenever an element of the framework is discussed in the body of the dissertation the element will be highlighted with bold text. Thus, when reading CoGScience related discussion, any words in bold in the body of the text should also exist in the framework diagram in Figure 8. For example, when we talk about a framework **process** such as the **coordinate process**, the actual framework elements will be displayed in bold (typically only once per sentence as they are here).

It should also be pointed out that although we attempt to make the framework conceptually complete, it is not possible to exhaustively represent all possible elements. We therefore assume (and expect) that other researchers will add components to this framework. For example, although **haptic sensory streams** are noted in our framework, the framework does not represent haptic sensory streams in detail (there are no **technology characteristics** associated with haptics). This is a decision of research scope, not a judgement of whether haptics are useful in artifact-centric collaboration. We assume that haptic researchers can and will refine the **haptic sensory stream** component of the framework to include a set of **technology characteristics** that apply to haptics. It is our belief the framework not only supports, but encourages such use.

We begin our discussion of the framework by splitting the communication process into two main domains; the **task domain** and the **technology domain** (see Figure 8). From a communication model perspective (Section 2.2.1.1), the boundary between these domains (the vertical dashed line in Figure 8) is where information (the signal) that needs to be communicated to accomplish a task (the effect) is encoded into one or more sensory streams (the medium) and transmitted to remote collaborators (receiver). This straight forward encoding of a basic model of communication (similar to that of Lasswell [Las48] or Shannon/Weaver [SW49]) is a fundamentally important process, and is in fact where the CREW and ETNA [MJA+02] frameworks are incomplete. That is, neither of these frameworks provides a mapping from a **human channel** for communication to a technology mediated **sensory stream** of information. Thus, this barrier between the task

and technology domains is explicitly encoded in our framework. In reality, the communication model that is the closest to our overall framework are the more complex models such as the Berlo Source, Message, Channel, Receiver (SMCR) model [Ber60, Section 2.2.1.1]. Such models not only include the signal, the medium, and the receiver, but also incorporate factors such as skills, attitudes, knowledge, and the environment into the communication process.

4.4.1 The Task Domain

The **task domain** of our framework is designed to help determine the desired effect of a communication task as well as the human communication channels that might be used to achieve that effect. We divide the task domain into four task based levels: **task, function, process,** and **channel** (see the top of Figure 8). Each level decomposes the communication process to a lower level of detail, making it relatively simple for a researcher to consider a task at varying levels of detail in the context of the framework.

4.4.1.1 The Task Level

The **task level** encodes McGrath's typology of tasks [McG93] and provides a high-level task decomposition (see Section 2.3.4.1 or Figure 4, p. 49). We maintain McGrath's four task quadrants, considering tasks that **generate**, **choose**, **negotiate**, and **execute**. McGrath's **generate** quadrant is divided into tasks that generate plans (**planning**) and generate ideas (**creativity**). The **choose** quadrant is divided into tasks where the goal is to reach the right answer (**intellective**) and tasks where the goal is to reach consensus on a preferred alternative (**decision**). McGrath's third quadrant, **negotiate**, extends the choose quadrant with an added intra-group conflict component. This quadrant is divided into tasks where the conflict is between different viewpoints within the group (**viewpoint**) and conflict is between different interests or motives (**interest**) within the group. McGrath's fourth quadrant considers tasks where the goal is to **execute** physical tasks. This quadrant is divided into tasks in which the group is in competition with another entity and the outcome will be either a win or a loss (**competition**) and tasks where the goal is to perform the task as well as possible (**performance**).

Unlike McGrath, our framework does not assume that a task cleanly fits in a single "task bin" exclusive of all other bins [McG93, p 165]. This view is too simplistic when

considering a real world task, as most tasks have features from more than one of the task categories. The task level decomposition helps to categorize a task, but we assume that many tasks will span two or more task categories. In fact, the CoGScience Framework's **task characteristics** (discussed below) partially overlap with McGrath's division between task quadrants and task types. We therefore encode McGrath's eight task types, but also list task characteristics that allow a researcher to consider other variables that may refine a task to span one or more of McGrath's task types.

4.4.1.2 The Function Level

The **function level** of the CoGScience Framework considers communication functions that one might carry out to accomplish a task. These functions are partially derived from McGrath's more detailed description of his typology of tasks and partially from the task domains that are used in the ETNA taxonomy [MJA+02]. The CoGScience **functions** also incorporate high-level task functions from Card *et al.*'s operations [Car08] (e.g. **search**, perform (**execute**), distribute (**report**), **describe**, extract (**generate**)), Isenberg *et al.*'s processes (e.g. browse (**explore**), **parse**, **discuss**, strategize (**plan**), clarify (**discuss**), select (**choose**), operate (**execute/manipulate**), validate (**decide**)) [ITC08], and Scott *et al*'s activities related to territoriality (e.g. **compare**, **search, sort, arrange**) [SC10].

One can think of functions as actions that need to be performed to accomplish a task. For example, for **choosing** and **negotiation** tasks, it is necessary to **discuss**, **negotiate**, **express ideas**, and **mediate**. For **execution** tasks, it might be necessary to **manipulate** data, **solve** problems, and **generate** ideas. In many ways the task and function levels are tightly coupled, in that the actions that need to be performed to accomplish a task in some ways define the task. This is why there is both a **negotiate task** and a **negotiate function**. The differentiating factor is that the task level classifies the task while the function level lists actions that are required to perform the task. Note that there is rarely a one-to-one mapping from task to function, and it is this mapping that defines the need to have both task and function levels in the framework. For example, the completion of a **competition task** may require a **negotiation function** to succeed.

4.4.1.3 The Process Level

The **process level** is slightly more fine grained, and considers the processes that one would utilize to carry out the task **functions**. This level can be thought of as distilling Gutwin and Greenberg's mechanics of collaboration [GG00] and workspace awareness work [GG02] with some aspects of Card's operations [Car08] and the **technology characteristics** from the CREW Framework [OO97].

Although the CREW Framework briefly discusses what we consider **technology characteristics** (**fidelity**, **clarity**, **delay**, etc), their technology section spends a significant amount of discussing what we call group **processes** (**coordination** and **awareness)**. We are very careful to create a separation between **functions**, **processes**, **channels**, and **technology**. For example, to **express** ideas (a **function**), one needs to **verbalize** and **engage** the audience (**processes**). To **manipulate** an object (a **function**), one needs to have workspace **awareness** and communicate **intention** (**processes**). Unlike the CREW Framework, at the process level, we do not consider **technology** and **technology characteristics** at all, but instead leave that consideration to the **technology domain** of the framework. Similarly, the CoGScience representation of Card's operations [Car08] are spread across multiple levels, as either **functions** (see previous section) or **processes**. For example, we represent Card's **search** operation as a function (it is a function that a group carries out to accomplish a task) while we represent Card's **monitor** operation as a process (it is a process that is important to a range of functions).

The CoGScience Framework suggests an extensive list of **processes**, including the ability to **verbalize**, **coordinate** (turn-taking and coordinating action**)**, communicate **intent**, control **access**, **engage** others, **monitor** others, **assist** others, **manipulate** artifacts, **modify** artifacts, and **create** artifacts. Like many aspects of the framework, we do not presume to have tabulated an exhaustive list of processes, but instead provide a fundamental base on which researchers can incorporate processes that may apply to the task they are considering.

It is important to note that in many senses, these processes are independent of whether the process applies to **conversation** or an **artifact**. The ability to make this differentiation (as suggested in [OO97]) is critical. For example, the communication **channels** and **sensory streams** (discussed below) required to support the **coordination process** are

very different if considering **coordination** (turn-taking) in the context of **conversation** (where **verbalization**, **body language** and **facial expression** may be important) and **coordination** in the context of an **artifact** (where **workspace awareness** and **gestural interaction** may be more important). Thus, processes also need to take into account whether they are applied to the conversation or an artifact. This feature is also encoded in the framework.

4.4.1.4 The Channel Level

At the most fine-grained level of our task decomposition, the **channel level** considers the human communication channels that we utilize to communicate. These human communication channels are derived from the related social psychology research in verbal and non-verbal communication. We utilize a categorization of communication channels that is distilled from Clark's research in language use [Cla96], McNeill's [McN92] and Kendon's [Ken04] research in gesture, and CSCW research of Bekker *et al.* [BOO95], Tang *et al.* [TL88], and Gutwin *et al.* [GG02]. This channel decomposition is somewhat similar to the Symbol Sets as suggested by Media Synchronicity Theory [DFV08], with extensions to consider the communication channels more explicitly. CoGScience channels are roughly broken down into **aural**, **gestural**, and other **non-verbal** channels [Cla96]. Aural channels are either **verbal** or **paralinguistic** channels (pitch, volume, intonation, rhythm, and emphasis). There are a number of ways to classify gestural communication [McN92][Ken04][BOO95][TL88]. Given that our focus is on artifact-centric collaboration, we incorporate artifact **manipulation** gestures and **pointing** gestures in the framework directly. We also include **spatial** gesture (gestures that indicate spatial relationships), **kinetic** gesture (gesture that acts out an idea), and **rhythmic** gesture (gesture that is timed with the discourse). Additionally, we list other important non-verbal communication channels, such as **body language**, **facial expression**, **eye contact** (ability to tell if individuals are looking at each other), **gaze awareness** (ability to tell where in the environment someone is looking), and **workspace awareness** (ability to tell how individuals are working within a workspace). We incorporate the work of Gutwin *et al.* [GG02] on workspace awareness to provide a set of **workspace awareness channels** such as the channels that communicate information about **location**, **control**, **alignment**, and **access**. These **channels** are listed in Figure 8.

As with task **processes**, task **channels** can also be divided along an orthogonal dimension. That is, most of the communication channels listed above can either be a result of **explicit** communication (information that is communicated with intent) or **consequential** communication (information that is communicate unintentionally) [GG00]. The framework captures these dimensions, suggesting that when a **channel** is considered in terms of the **functions** and **processes** it provides it should also be considered whether it provides that communication in an **explicit** or **consequential** form. For example, consider the task **channels** in a remote application that uses a Smartboard to manipulate a shared tele-pointer. This provides an **explicit** gestural **pointing** channel, as one intentionally touches the screen to obtain a specific outcome (moving the pointer). It does not provide any **consequential pointing** cues that capture when a user approaches the Smartboard with the intent of carrying out an operation but does not complete the operation for some reason (they were interrupted, someone else performed an action). Without differentiating between these types of channels, it is difficult to capture the subtleties of the **pointing gesture** in this instance. The **explicit** channel is important to **processes** such as **manipulating**, **modifying**, and **creating** artifacts while the **consequential** channel is important for **coordination**, communicating **intent**, and **monitoring**.

4.4.1.5 Communication Characteristics

We also encapsulate the task level **communication characteristics** that affect the group communication process as part of the framework. These characteristics are orthogonal to the task levels (can be applied to any level) but are still part of the task domain of the framework. Many of these characteristics are derived from other related frameworks; in particular those presented in the CREW Framework [OO01], McGrath's Typology of Tasks [McG93], and Daft and Lengel's Media Richness Theory [DL86]. These characteristics also encompass many of the contextual challenges (group size, group member familiarity, group member backgrounds, familiarity with the task, duration of the task, and group attitude) listed by Isenberg *et al.* in their analysis of collaborative information visualization on digital tabletops [IHH+10] as well as Kirk *et al.*'s levels of prior grounding in the use of remote gesture in object-specific collaboration (level of experience, novelty of the task, and urgency of the task) [KRF07]

We present our framework's communication characteristics from three perspectives: **characteristics of the task**, **characteristics of the environment** in which the task is taking place, and **characteristics of the group** carrying out the task. The framework **task characteristics** consider things like the **nature of the material** (abstract ideas or concrete objects), the **task coupling** (loosely or tightly coupled), **familiarity** with the task, **duration** of the task, and whether the task is **ambiguous**, **uncertain**, **creative**, **exploratory**, **difficult**, or **complex**. Note that these characteristics help to classify a task within the **task level** of the CoGScience Framework.

Characteristics of the **environment** or situation that are considered are **organizational norms** (reward structure, work norms, and organizational routines), **time of day** the task takes place, and whether there is **competitiveness**, **urgency**, importance, **conflict** or **emotion** around the task. Characteristics of the **group** are **familiarity** (both of individuals in the group as well as with the group as an entity), **size** of the group, **composition** of the group (homogeneity of abilities, homogeneity of seniority, cohesiveness, and trust), **individual** characteristics (skill, personality, and motivation), and group **distribution** (how the group is distributed across multiple sites). Although these characteristics are most applicable at the **task** and **function** levels, it is useful to consider them at both the **process** and the **channel** levels of the framework as well. For example, whether or not the group has to work together closely to solve an artifact-centric problem (the **task coupling**) has an impact on the importance of the **gestural channels** for that task.

4.4.2 The Technology Domain

In the previous section, we considered the **task domain** of the CoGScience Framework. In this section, we explore the **technology domain**. We divide the **technology domain** into two parts, the **sensory streams** that are used to transmit the task domain communication channels and the **technology characteristics** of those sensory streams. Sensory streams in this case are communication streams that use a sensual modality (**visual**, **auditory**, **haptic**, **taste**, and **smell**) to communicate information. These are the fundamental building blocks of how information is communicated in technology mediated collaboration environments (and indeed in face-to-face communication), and therefore all information that is communicated to a remote collaborator will utilize one or

more such streams. **Visual** and **auditory streams** are by far the most common in distributed collaboration, but haptic feedback has been used in some research. Our research focuses on visual and aural sensory streams, but as discussed in Section 4.4, there is no reason why the framework cannot be extended to incorporate a **haptic sensory stream**.

Like in the **task domain**, the CoGScience Framework encapsulates the **technology characteristics** that are applicable to the **aural** and **visual sensory streams**. Unlike the technology characteristics presented in the CREW [OO97] Framework, the technology characteristics in our framework are completely separate from the task domain. We consider the mapping from **task** (and in particular task **process** and **channel**) to **visual streams** as one of the most important steps in analyzing a distributed, artifact-centric system.

Both the CREW Framework and ETNA Taxonomy merge the **task** and **technology** domains. For example, the presentation of the CREW Framework in [0097] discusses technology characteristics in the context of particular processes and channels, but does not enumerate a concise set of technology characteristics that can be applied to such processes and channels. Similarly, the ETNA taxonomy applies different technology characteristics to technologies that are used to represent personal space (telepresence) and task space (teledata). In addition, both CREW and ETNA have the technology characteristics intimately associated with the division between personal space (technologies to support the conversation in CREW and telepresence in ETNA) and task space (technologies to support the work object in CREW and teledata in ETNA). The CoGScience Framework uses the **task process** abstraction level to differentiate whether communication processes are used to support the **conversation** (conversation /telepresence) or the **artifact** (work object/teledata). This removes all task specific considerations from our **technology characteristics**, making what we believe to be a much cleaner process in considering these characteristics for a task specific set of **processes** and **channels**.

We distil the **technology characteristics** in CoGScience from those used in the CREW framework and the ETNA taxonomy, with a concentration of removing task considerations from the list of technology characteristics. We specify a more extensive

set of technology characteristics than ETNA, basing these extensions on our extensive review of the collocated and distributed collaboration literature (see Section 2.3). In particular, unlike ETNA, the CoGScience Framework suggests that technology characteristics should be considered consistently for all **task processes** and **channels**.

The CoGScience audio characteristics are broken down into:

- **fidelity** (quality of the audio channel, including sampling rate);
- **delay** (latency between the speaker speaking and the receiver hearing audio);
- **echo** (the quality of the echo cancellation);
- **control** (whether individuals can control their own or other's audio);
- **spatiality** (mono, stereo, or spatial sound);
- **interactivity** (level of interaction among parties); and
- **reciprocity** (whether all participants have equal level of quality)

Similarly, video characteristics are broken down into:

- **fidelity** (resolution);
- **quality** (impacts of compression in color, space, or time);
- **temporality** (how fast the source video changes);
- **frame rate** (how fast the video is encoded and decoded);
- **clarity** (level of focus and depth of field);
- **field of view** (what is shown on camera – people or work object);
- **delay** (latency between when an action occurs and when a viewer observes it);
- **lighting** (quality and amount of lighting);
- **control** (whether participants can control their own and other's video);
- **interactivity** (level of interaction of participants);
- **reciprocity** (whether all participants have an equal level of quality);
- **size** (size of the physical display of the visual channel);
- **orientation** (orientation of the display (horizontal/vertical) of the channel); and
- **spatiality** (spatial arrangement of the users around the display).

The approach taken to define the technical characteristics in the CoGScience Framework makes the division between task and technology quite clear. This clarity

between task and technology is one of the key contributions of the CoGScience Framework.

4.4.3 Measures and Outcomes

Our framework makes use of a set of **measures and outcomes** that help us to understand the effect of the many parameters that exist in this framework. From an experimental standpoint, these can be thought of as the dependent variables that one might measure experimentally to determine outcomes (while controlling for the other variables in the framework). The framework utilizes measures of **process**, **task**, **group**, and **cognition**. Our task and group outcomes are modelled after the CREW Framework's measures [OO97]. We extend these measures to include measures of the collaboration process as well as measures of the cognitive process.

Measures of **process** include measures for **task process** (**depth/breadth of analysis**, structure of the work, and **efficiency**), **communication process** (amount of clarification, **turn-taking**, non-verbal communication, **gestures**, **digression**, and socialization), and **inter-personal process** (amount of **conflict**, affect/mood, and **participation**). **Measures of task** include those of **task outcome** (quality of the work, time to perform task, and cost to complete task) and **satisfaction** of the stakeholders (**individuals**, **group**, **organization**, and **client**) with the task outcome. **Measures of the group** include measures of attitude towards **group work** (satisfaction with the process and group buy-in) and attitude towards the **group** (would they work together again?). Lastly, we also include **measures of cognitive process**, primarily focussing on measures of **extraneous cognitive load** (cognitive load induced by the presentation of material), **intrinsic cognitive load** (cognitive load induced by the intrinsic complexity of the material), **attention,** and **understanding**.

4.4.4 CoGScience Summary

Although the CoGScience Framework suggests many incremental changes to existing frameworks, the key contribution of this framework is the structure of the task and technology domains. The CoGScience Framework extends existing frameworks with better defined task and technology domains, while at the same time providing a mechanism to bridge this gap. The ability to map collaboration **task** to collaboration

function, to collaboration **process**, to communication **channel**, and finally to **technology** implementation, is a new capability in CSCW frameworks.

4.5 Using the CoGScience Framework

One of the key goals of the CoGScience Framework is to provide a tool for researchers, designers, or software implementers to rigorously explore artifact-centric collaboration. Like Olson, we are interested in "Making Sense of the Findings" [OO97], with the goal of this framework to help establish a common grounding in which researchers can consider a wide range of distributed, artifact-centric collaboration research. One of the key differences between the CoGScience Framework and the related frameworks on which it builds is the number of elements encapsulated within the framework. This is intentional, as the application of other more succinct frameworks (see Section 2.3.4 and Section 4.3 for details on these frameworks) did not capture enough detail about the collaboration processes we encountered when building CoTable. As discussed in Section 4.4, we take a similar approach to Dennis *et al.* [DFV08] and Shneiderman [Shn96] in that we believe it is necessary to refine the granularity of the task and technology aspects of remote collaboration beyond a small set of task categorizations. CoGScience therefore errs on the side of too much detail, rather than too little. The goal of the framework is not to use every element in every situation, but to use the CoGScience framework to direct the researcher or developer towards asking appropriate questions about the categorization of the collaboration task being considered. Below, we briefly explore two ways of applying the CoGScience Framework to related research in this area.

4.5.1 A top-down approach

The top down approach entails looking at a collaboration driven by task. That is, by drilling down through the **task domain**, the framework directs the researcher towards a set of questions that will pull out important task requirements. For example, what is the high-level type of the task being undertaken (e.g. a **creative** task – brain storming), what are the **functions** required (**express ideas**, **generate ideas**), what are the **processes** used to carry out those **functions** (**verbalize**, **workspace** use, **coordinate** actions), and finally what communication **channels** are used (**verbalize**, **gesture**, etc). Once the appropriate communication **channels** are determined, it is possible to consider the **technologies** (and

the relevant technology **characteristics**) that would be useful to create a collaboration environment that would meet the collaboration needs of the task.

Orthogonal to the task categorization, the framework also suggests to the researcher a set of **task**, **environmental**, and **group characteristics** that might affect the group task. What is the **size** of the group? Do they **know** each other? Is the task **difficult**? Is it **urgent**? How **long** does a brainstorming session take? All of these questions flow out of the framework naturally.

If a researcher is considering an experimental study of such a task, the framework leads him/her to a set of **measures** that might be quantified in an experiment (**efficiency** of the process, **effectiveness** of turn-taking, amount of **digression**, number of **gestures**, or **attention** paid to artifacts). It also helps the researcher to identify the variables that the experiment might want to manipulate and control (different **fidelity aural stream**, ability to see a **visual gestural stream**). Again, these suggestions flow naturally from the framework.

4.5.2 A bottom up approach

The bottom up approach to using the framework considers group work from the communication and technology perspective. Consider the case where one is analyzing a currently existing collaboration. How does one analyze the tools currently used? The framework suggests decomposing the collaboration into **sensory channels**. What **visual channels** are being used (video feed of participants or visual shared workspace)? What **aural channels**? What are the characteristics of those channels (**fidelity**, **field of view**, or **synchronization**)? How do those characteristics impact the functionality of the task?

Again, using the example of an experimental study, one might want to measure the effect of changing the video **quality** and **fidelity** of a task space **visual stream** (e.g. different video compression implementations and video resolutions respectively – our independent variables) on the attention paid to artifacts (a **cognitive** measure – our dependent variable) in a distributed collaboration. The most daunting thing about this example is the sheer number of other parameters that exist in the CoGScience Framework that could potentially impact our measure. When designing such an experiment it is desirable that all other **characteristics** (**task** and **technology**) be controlled. Recall that in Section 2.3.2.1, we saw that research studies have shown mixed results on the value of

video in remote collaboration. Considering the complexity of this domain, as captured in the CoGScience Framework, one can perhaps see why this might be so. Clearly video is not important all the time, but in which situations and why is still an open research question.

Perhaps the most useful application of the CoGScience Framework is in the exploration of where the framework **task** and **technology** domains intersect. Recall from our discussion of communication models that the boundary between task and technology is where information is encoded and decoded to achieve a specific effect. One of the key questions that the framework can help to answer is, given that we have identified a set of human communication **channels** that are required for a specific task, how are those channels being encoded and sent via the **sensory streams** to the remote participants? How are the remote participants decoding those sensory streams? For example, if **facial expression** is a required communication channel for the task, is it being encoded and sent as a **sensory stream**? If so, is the sensory stream achieving the desired communication effect?

The top-down and bottom-up example processes given above are necessarily short. The CoGScience Framework is complex, and describing its application in detail takes a significant amount of space. As the CoGScience Framework is a fundamental component of this research, we apply the framework in various roles throughout this dissertation. More details on this application can be found in later chapters.

4.6 Conclusions

In this chapter, we present a framework for analyzing distributed, artifact-centric collaboration. The framework builds on past work in this area, extending previous frameworks on several dimensions. We expand on the composition of the **task domain** used in most frameworks, utilizing four levels: **task**, **function**, **process**, and **channel**. This decomposition of task helps the researcher go from a high-level **task** description to the specific communication **channels** required to accomplish that task. We also present a list of **task characteristics** (**task**, **environment**, and **group** characteristics) that potentially affect this communication process. We present a detailed decomposition of the **technology domain** with our framework, listing the physiological **sensory streams**

that we use to process information as well as the **characteristics** that apply to these sensory streams when communicated using technology.

One of the foundations of the framework is its ability to explore the intersection between the **task domain** and the **technology domain**. It allows researchers to explore how a collaboration environment maps the task level communication **channels** to technology oriented **sensory streams**. We believe this mapping from the task domain to the technology domain is a critical component in understanding complex collaboration environments and we have therefore encoded this directly in the framework.

Perhaps the most useful aspect of the framework is the ability to utilize it to synthesize a structured, coherent view of a collaboration task from the range of sources of information that are available to us. The framework provides a detailed view that encompasses the social sciences (communication, social psychology, and cognitive psychology) and computer science (HCI and CSCW). The framework is strongly grounded in past research on collaboration frameworks, but also makes extensive use of the experience and knowledge gathered through the research presented in this dissertation.

Note that in this dissertation, the CoGScience framework is used to study artifact-centric scientific collaboration. It is our belief that the framework would be a useful tool for the study of any artifact-centric collaboration, with the caveat that some of the task and technology characteristics may need to be refined to apply to other application domains. Throughout this dissertation, we demonstrate the utility of the framework by applying it to the domain of distributed, artifact-centric scientific collaboration in the following ways:

- We apply the framework in an analysis of the CoTable system (Chapter 5);
- We use the framework to design, implement, and synthesize the results of the research studies (both qualitative and quantitative) presented in this dissertation (Chapter 6 through Chapter 11); and
- We use the framework to suggest the Collaborative Group Science (CoGScience) design guidelines (Chapter 12).

Part III - Studies

5 Distributed Tabletop Collaboration (CoTable) – A Case Study

Co-located tabletop collaboration embodies one of the most compelling and richest technology-mediated face-to-face environments [SGM03], utilizing a physical environment that invites cooperative work with digital data. Although the seminal research on the impact of distance on rich, collocated interaction environments took place in the late 1980s and early 1990s [TL88, TM90] there has been a recent renewal of interest in this area [TPI+10] (see Section 2.3.3 for details of this research). This is particularly relevant today, as touch sensitive devices that use gestural interaction are becoming commonplace, both at the large scale (e.g. Smartboard [Smart]) and the small scale (e.g. iPhone), and may soon be ubiquitous.

Two of the primary research objectives of this research are to:

- *Objective 2:* Develop a deep understanding of how scientific researchers interact with digital artifacts when they collaborate.
- *Objective 3: Evaluate advanced collaboration modalities and technologies for scientific* collaboration.

In order to explore these domains, a distributed, tabletop collaboration environment (CoTable) was created. The goal of this research was to explore the issues and problems that arise when a natural interaction environment, such as that presented to users when communicating over a tabletop, is altered by having one or more of the users communicating from a remote location. The tabletop environment was chosen because it was a natural, multi-user environment that is rich in subtle communication channels. It is these communication channels that are to be explored, with the goal of identifying mechanisms by which they can be reproduced or replaced to allow the distributed user to communicate effectively with the other collaborators in the group.

In particular, this chapter helps to answer a number of our key research questions, providing valuable insights into *the role that digital artifacts play in collaboration*, *what information is lost when artifact-centric collaboration takes place at a distance*, and *what communication channels can be used to encode information for distributed, artifact-centric collaboration*. In addition, this chapter allows us assess to *assess how researchers use advanced collaboration technologies* and *how well those technologies work*.

We proceed as follows. Section 5.1 outlines the CoTable hardware and the VideoBench video editing software. In Section 5.2, we apply the CoGScience Framework to both the task and technology of video editing using CoTable and VideoBench. In Section 5.3 we discuss the usage scenarios that contribute to our analysis, in Section 5.4 we analyze this usage in the context of our CoGScience analysis, and in Section 5.5 we summarize our findings.

5.1 CoTable and VideoBench

The goal of the CoTable environment was to provide us with a technology prototype with which we could consider a range of advanced collaboration scenarios. We gain experience with CoTable through carrying out a case study of a distributed video editing application (VideoBench) implemented on the CoTable system. As discussed in Section 3.3, the technology environment at which we are targeting our collaborative tools is a sophisticated one. Although the CoTable system is a technology prototype, it is not fundamentally different than those systems considered later in this dissertation. We assume the use of touch-screen interaction, multiple displays, multiple cameras, and a high-fidelity audio system such as those described in Section 3.3 and studied in Chapter 6 and Chapter 7.

Our initial approach to creating the CoTable system was to use existing CSCW and group work frameworks to analyze the collaboration needs of computer mediated tabletop collaboration environments. In particular, we used the Mechanics of Collaboration [GG00], the ETNA Taxonomy [AMJ+02], and the CREW Framework [OTC+02] (see Section 2.3.4 for details on these frameworks). The intent was to use these frameworks to do the following:

- Tabulate the collaboration needs of collocated artifact-centric collaboration;
- Identify the information that is lost when users are not physically collocated;
- Determine and implement mechanisms that provide some (or all) of the information that is lost to the distributed users; and
- Analyze the use of the implementation.

It was through the application of these frameworks to the CoTable system that we identified several gaps in these frameworks. This led directly to the creation of the

CoGScience Framework (Chapter 4). As a result, in the remainder of this chapter we do not consider these frameworks any further. Instead, we utilize the CoGScience Framework in our analysis of the CoTable environment. See Section 4.3 for a detailed description of how these frameworks contributed to the creation of the CoGScience Framework.

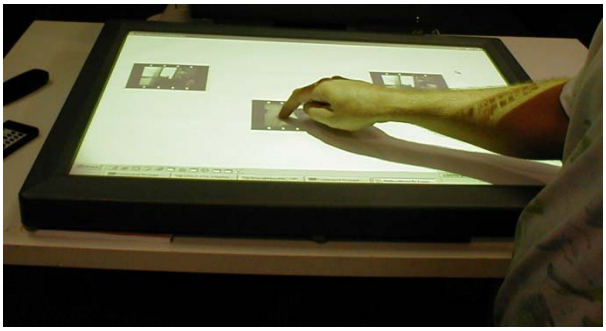


Figure 9: The collocated CoTable system.

5.1.1 The CoTable System

The collocated CoTable system consists of two main components, a touch sensitive tabletop device and an overhead projector that projects an image on the tabletop. The Mitsubishi DiamondTouch (DT) table [DL01] provides gestural input for the system. The DT table is a multi-user, multi-touch input device that can detect a large number of contact points and can associate each contact point with an individual user. Through the processing of touch information over time, the DT table can detect gestural interaction as users perform actions on the table.

The application display of the CoTable system is where artifacts are displayed. The system uses the DT as a top projected display surface. A projector is mounted above the table, illuminating the DT with an image from a computer application. One of the features of this system is that it allows the user's gestures to be collocated with the artifact (as illustrated in Figure 9).

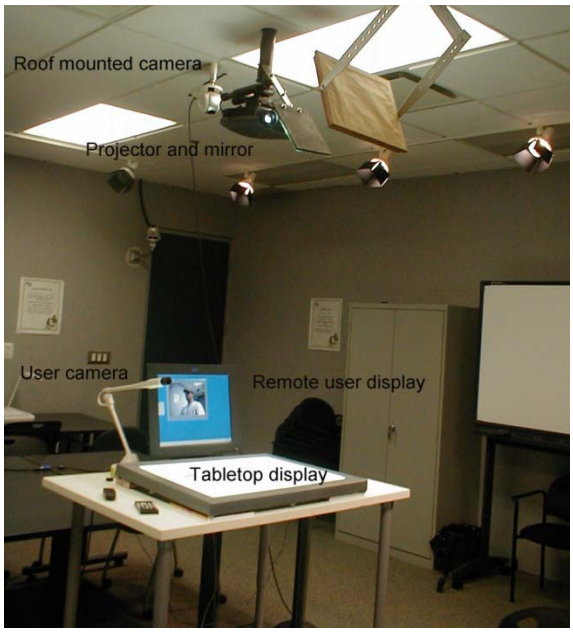


Figure 10: The distributed CoTable system

5.1.2 The Distributed CoTable System

The distributed co-table system extends the collocated system with the ability to interact with a remote user. These extensions consist of the following:

- A standard LCD monitor located across the table from the user. This provides a mechanism to display information sent from the remote user to the tabletop user.
- A camera next to the LCD monitor. This provides a mechanism to capture and send a view of the tabletop user to the remote user.
- A camera mounted on the ceiling. This provides a mechanism to capture and send a view of the tabletop and the actions that are performed over the tabletop to the remote user.

A view of the technical components of the distributed CoTable system can be seen in Figure 10.

Figure 11: The distributed remote desktop configuration

Finally, because we did not have access to two DT table systems during our case study, it was necessary to implement a simpler environment for the remote user. The remote user system consisted of a single LCD monitor, a single camera, and a mouse and keyboard for interaction. The application display that is shown on the DT table is displayed on the monitor. The camera views from the CoTable system (the user camera and the tabletop camera) are also displayed on the LCD monitor. The camera is used to capture and send video of the user to the CoTable system and the mouse and keyboard are used to interact with the application. This simulates a hub and satellite model of working [VTC+10], where the main hub of collaborators is collocated (the tabletop) and there are one or more satellite individuals (the desktop) joining from a remote site. Note that this hub and satellite model of working is also common in our ethnography in Chapter 7. A photo of the remote user station can be seen in Figure 11.

5.1.3 The VideoBench Application

As described in Section 1.1.1, the focus of this research changed from artifact-centric collaboration in general to the specific domain of artifact-centric, scientific collaboration *after* our initial experiments with the CoTable system. Our experiences with CoTable did not revolve around scientific collaboration, but instead considered another common artifact-centric task – that of video editing. Video editing was chosen as an artifact-centric task on CoTable for two primary reasons. First, it is a visually complex artifact-centric task (manipulating and editing video segments) and requires high fidelity control

(control is required at frame level for editing videos). Second, video editing, like photo-editing [SLV+02], is a social activity that is conducive to tabletop interaction.

Our implementation of a tabletop video editing application is called VideoBench, an interactive application for the manipulation and editing of video clips (see Appendix 15.2 for a detailed description of the VideoBench application). The initial collocated application was developed by a group of students at the University of Victoria to explore gestural interaction on a tabletop device [CCG+03]. The application supported collocated collaboration and used gesture as an interaction mechanism. It allows multiple users to play (play, pause, rewind, fast forward, stop) and edit (cut, splice, resize) a set of video clips using gestures on the DT table.

It is worth noting that these are HCI gestures (communicating to the computer) rather than HHI gestures (communicating between people). Our primary interest in this research is in HHI gestures (how people communicate with each other at a distance using gestures) not the HCI aspect of how gestures are used to interact with VideoBench software. The gesture set that is supported in the VideoBench application is important, because it enables natural, gesture-based interaction with digital artifacts. At the same time, it is not the focus of the research presented here. Although we have carried out a study of how gestures were used in the collocated version of VideoBench [CFM+03], that research is not discussed in this dissertation.

5.1.4 The Distributed VideoBench Application

We extended the original face-to-face VideoBench application to create a new distributed version of VideoBench. The distributed application functions like the collocated version except that it supports the application running at two or more physically disjoint locations. The applications exchange state such that when a user at one site performs an operation, the user at the remote site sees the operation occur as if it was performed locally. The user's gestures on the tabletop are communicated to the user at the remote site through the drawing of an icon (in this case, a circle of roughly fingertip size) that represents where the remote user is touching the table. Up to two contact points can be communicated per user with each touch point leaving behind a gestural trace as the user interacts with the table. More details of the implementation of the distributed application can be found in Appendix 15.2.2 and in [Cor03].

5.2 Applying the CoGScience Framework

In this section, we utilize the CoGScience Framework to analyze the VideoBench software and how it is used on the CoTable system. We proceed as follows:

- We use the CoGScience Framework in a top-down fashion to analyze collocated video editing as a general collaboration task (Section 5.2.1).
- We use the CoGScience Framework in a bottom up fashion to analyze how communication channels are communicated using technology in the specific case of the VideoBench software deployed on the CoTable hardware (Section 5.2.2).

This analysis demonstrates the strengths of the CoGScience Framework in both task driven and technology driven scenarios. In Section 5.4, we explore the intersection of these two analyses in the context of the experiences the author and two other users had while testing and using the distributed VideoBench system.

5.2.1 CoGScience: Studying Collocated Video Editing as a Task

In order to understand the specific needs of video editing, we consider a simple, collaborative video editing task. Two users are working together using VideoBench, with the application initialized with two video clips in the system. The task consists of two steps. First, collaborators are required to choose one of the videos, split the video at a location of their choosing, insert the second video clip between the two new video clips, and join them together into a single video. The second step is to take the resulting single video, split the video into three clips at the locations where they were originally spliced together, and recreate the two original videos.

This task requires communication so the operations between the two users can be coordinated. The task has sections where serial steps have to be taken as well as sections where users can perform parallel tasks. The first part of the task is cooperative and creative (collaborators need to agree on a video and where to split it) while the second part of the task is more structured and mechanical (requiring the collaborators to split the videos at a specific location and restore the two videos to their original format).

5.2.1.1 CoGScience: The Task Domain and Video Editing

In order to better understand the collaboration needs of such a video editing task we utilize the four levels of the **task domain** of the CoGScience Framework. This equates to

the top-down approach of applying the CoGScience Framework as described in Section 4.5.1. As usual, when applying the CoGScience Framework, all framework components are listed in bold, allowing the reader to refer to the CoGScience Framework diagram (Figure 8, Page 83) as the framework is applied.

At the **task level**, the video editing task has two phases. The first phase is both a **creative task** (ideas are generated) and a **decision making task** (decisions about which video to split and where). The second phase is a **performance task**, where the goal is to achieve a specific outcome (performing precise and specific video editing operations).

From a **task function** perspective, fundamentally all aspects of the task require the ability to **manipulate** artifacts. **Creativity tasks** require the ability to **generate ideas**, **decision making** tasks require the ability to **discuss** and reach **consensus**, and **performance tasks** require the ability to **execute**.

From a **task process** perspective, many of the processes in the CoGScience Framework are applicable to the tasks being considered here. These include:

- **verbalizing, turn-taking** (who is going to perform the next video editing action);
- **communicating intention** (determining whether someone intends on performing a video editing action. E.g. reaching towards a video);
- **coordinating action** (coordinating the video editing actions that are taking place)**;**
- **awareness** (being aware that a collaborator is performing an action);
- **monitoring** (monitoring the progress of a video editing operation); and
- **assisting** (assisting a collaborator when they are having a problem performing an action).

Note that many of these processes apply to both the **work object** and the **conversation**.

Similarly, almost all **communication channels** represented in the CoGScience Framework (**verbal**, **paralinguistic**, **gesture**, **facial expression**, **body language**, **eye contact**, **gaze awareness**, **workspace awareness**, etc.) appear to be relevant to the task. It is important to point out that channels that support both **explicit communication** (communication that is intentional, such as **pointing**) and **consequential communication** (communication that is unintentional, such as **body language**) are required.

It is not terribly surprising that by the time we get to the **channel level**, all channels appear relevant to our task. Our high-level video editing task has been broken down into

three sub-types of tasks, **decision making**, **creative**, and **performance**. Given that the sub-tasks span three of the eight task types and three of the four task quadrants in the CoGScience Framework (recall that the CoGScience task level is based on McGrath's task circumplex [McG93] discussed in Section 4.4.1.1), the fact that many processes and channels are required is not surprising. In fact, in our experience with applying the CoGScience Framework to artifact-centric collaboration, most tasks require a wide range of processes and channels. In addition, we find that most collaborative applications also provide at least some level of capability in communicating most of these processes and channels. Thus we are rarely left with a binary question such as is process P or channel C required (when designing a system) or provided (when assessing an existing system). Instead, we are faced with the question of the degree to which process P or channel C is required/provided. The CoGScience Framework is particularly useful at helping to answer this question.

5.2.1.2 CoGScience: Communication Characteristics and Video Editing

Characteristic Type	Characteristic	Analysis
Task	nature of material	common media, complex operations
	creative, informal	yes
	difficulty, complexity	low
	duration	short
Environment	organization	social context
	competitive, urgency, conflict	low
	emotionality	possibly emotional
Group	size	two people
	familiarity (individual)	friends/family
	familiarity (group)	familiar
	composition	varied

Table 1: Video editing communication characteristics

We next consider the **communication characteristics** of the CoGScience Framework (Table 2). In terms of task characteristics, the **nature of the material** (video) is relatively simple but the application supports relatively complex operations on that material. The video editing task considered here is **creative** and **informal** with a low level of **difficulty** and low **complexity.** It is short in **duration**, with the time primarily determined by how long the collaborators take to decide which video to split and where. The **environment** is a social one, and there is no **competitiveness**, **urgency**, or **conflict**. **Emotionality** could

play a role in the video editing task, depending of whether the content of the video is of an emotional nature to the collaborators. The **group** is small in size (two in our case, but could be larger). In a social setting, the **individuals** would likely be **familiar** with each other but performing the video editing task as a **group** may not be **familiar**. The group **composition** can be highly varied, with members ranging in age, gender, and experience.

5.2.1.3 CoGScience: Video Editing Summary

Our application of the CoGScience Framework to the specific video editing task considered above demonstrates the utility of the framework. In considering the task and function levels of CoGScience we see that the task actually consists of two sub-tasks, and there are a relatively small set of functions that are required to accomplish these tasks. When considering communication processes and channels, CoGScience suggests that there are many processes and channels that are important to the video editing task. It is therefore necessary to consider the relative importance of all of these processes and channels and how they might be communicated to accomplish the task. In the next section, we use the CoGScience Framework to perform a bottom-up analysis of how VideoBench, implemented on the CoTable hardware, provides these communication channels.

5.2.2 CoGScience: Studying Distributed Video Editing using VideoBench

So what happens when we take a rich, artifact-centric collaboration task such as video editing and attempt to deliver this task using the distributed VideoBench application on the CoTable hardware? This system provides a number of sensory streams to remote collaborators. How can we gain an understanding of the information that is communicated by these streams? How do we analyze the effectiveness of the information that is sent? In this section, we consider this problem by applying the CoGScience Framework to the **technology domain** of the distributed video editing implementation embodied by CoTable and VideoBench. By exploring the technology domain of a specific implementation (CoTable and VideoBench), we are using CoGScience in a bottom-up analysis (see Section 4.5.2). That is, we analyze the technology characteristics of the aural and visual sensory streams that are provided by the CoTable and VideoBench implementations.

5.2.2.1 The aural stream in CoTable and VideoBench

In CoTable, remote collaborators use headphones with a microphone to provide an **aural sensory stream**. We use RAT [UCL] to transmit the audio between collaborators, providing relatively **high fidelity** (better than a typical phone line), full-duplex **interactivity** (both parties can talk at once), and monaural **spatiality** (non-stereo). The headphones and microphones result in no **echo**. Both the tabletop and the desktop systems use the same audio equipment, so the communication is **reciprocal**. Since the two systems are close to each other and connected by a high bandwidth switched (100 Mbps) network, the **delay** or latency is very low. The headphone level, microphone level, and audio mute are **controllable** at both endpoints.

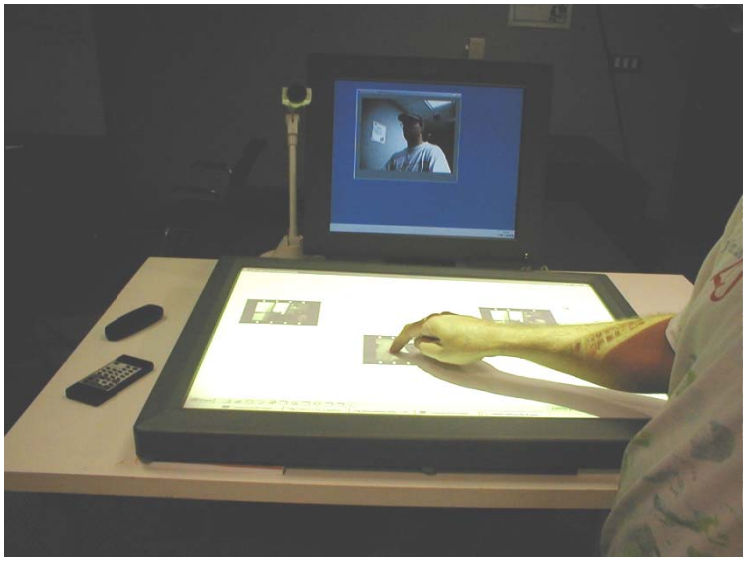


Figure 12: CoTable system in action

5.2.2.2 Visual sensory streams in CoTable and VideoBench

How visual sensory streams are utilized to meet the collaboration needs of this task is a complicated question. In fact, much of the disagreement about the value of visual communication channels in distance collaboration stems from exactly this question (see Section 2.3.2 for a detailed discussion). The CoGScience Framework was created to help resolve this problem, enabling the mapping of task to technology in a much more rigorous fashion. We tackle the problem by considering which communication channels are required to provide a useful personal space and task space**.**

In terms of providing a personal space, we attempt to mimic the “across the table” feel of a tabletop interaction. We provide a visual stream of the remote collaborators upper

torso (head and shoulders) and present this stream on an LCD monitor on the opposite side of the table (the **personal visual stream**). We use the VIC [MJ95] video tool to provide this video stream. The camera that captures a local collaborator's video feed is either on top of or next to the LCD screen on which the remote collaborator's video feed appears.

A **visual stream** of task space is provided through the shared application provided by VideoBench (the **application visual stream**). That is, the artifacts, and the actions that are performed on the artifacts, appear as an intrinsic part of the application. The application is displayed on the DT table in the tabletop configuration and on a normal LCD monitor in the remote station configuration. To a small degree, this visual stream provides a simple view of reference space as well, as gestures made by dragging an artifact on the table or dragging the fingers across the table allow action, and to some degree gesture, to be communicated.

Finally, a **visual stream** targeted at reference space (the intersection of personal and task space) is also provided (the **workspace visual stream**). This is done by mounting a camera on the ceiling, capturing a video of the tabletop, and sending the remote video stream to a collaborator using the VIC video tool. This provides a visual stream of both the VideoBench application and any physical actions made over the table by the tabletop user (reference space). Two implementations of the CoTable system are shown in Figure 10 (the implementation used in the case study) and Figure 13 (a non-experimental implementation created in one of the WestGrid collaboration rooms). A view of the CoTable system and the VideoBench software in action is shown in Figure 12.

We now consider the **technology characteristics** of the visual streams used in the CoTable/VideoBench system. It is important to note that the technology characteristics are specific to an instantiation of the CoTable system. That is, the characteristics that we tabulate below define this instantiation (e.g. we study the workspace visual stream at specific fidelity, quality, frame rate), and if we change a setting (e.g. resolution) of a specific characteristic (e.g. fidelity), we are considering a different system from a CoGScience perspective. In fact, it is often exactly such changes in the technology characteristics (e.g. different fidelity video) that are the independent variable in many experimental studies in the research literature.

Figure 13: A non-experimental CoTable implementation.

The technology characteristics of these three visual streams are listed in Table 2. Note that many of the characteristics of the **personal visual stream** and the **workspace visual stream** are similar, primarily because the same software tool is used to send the video to the remote collaborator. We consider these characteristics below, focussing on those characteristics that differ across the three visual streams.

	Visual Sensory Channel		
Characteristic	Personal	Workspace	Application
Fidelity	352x288 (CIF)	352x288 (CIF)	1024x768
Quality	Moderate (h261)	Moderate (h261)	High
Temporality	Moderate	Moderate	Low
Frame rate	24	24	10-15
Clarity	Good	Good	Good
Field of view	Head/Shoulders	Tabletop	Tabletop
Delay	Low	Low	Moderate
Lighting	Normal	Backlight	N/A
Control	No	No	Partial
Interactivity	High	High	High
Reciprocity	Yes	No	No

Table 2: Technology characteristics for CoTable/VideoBench

The personal and workspace visual streams are relatively low **fidelity** (353x288 pixels) and of moderate **quality** (the h261 codec implies compression and motion artifacts) while the application visual stream is high **fidelity** (1024x768 pixels) and high **quality** (no compression artifacts). **Temporality** of the visual field (how fast it changes) is moderate for the personal and workspace streams (as they capture human motion) but is relatively low for the application stream (interactions with the application are slower and more

deliberate). The **frame rate** for the personal and workspace streams is relatively high (24Hz) while we estimate that the application visual stream updates its state at between 10Hz and 15Hz. Both the workspace and application visual streams capture the same **field of view**, that of the tabletop. The workspace visual stream had a slightly wider **field of view** than the application visual stream (see Figure 14). The personal visual stream captures the head and shoulders of the collaborator.

Figure 14: CoTable top camera view

All visual streams have relatively low **latency/delay**. The application visual stream introduces **latency** in order to interpret HCI gestures, but the **delay** between an action taking place at a local site (e.g. dragging a video) and the result of that action being displayed at the remote site (e.g. the video moving) is relatively small. The room used in our testing had standard office **lighting**. The **lighting** for the workspace visual stream is an issue, as the brightness of the screen caused the user's arms to be quite dark compared to the desktop. This made skin tone and other subtle visual details difficult to discern. Although users had very little **control** over the visual streams, users could control whether gestures were communicated in the application visual stream by choosing to touch or not touch the table while performing a gesture. Since we only had one tabletop device, the level of **reciprocity** was compromised for both the workspace and application visual streams. The desktop user received all of the visual streams sent from the tabletop user. Since the desktop user did not have a tabletop system, there is no reference space visual stream (the overhead camera view) sent from the desktop user to the tabletop user.

5.2.2.3 VideoBench and CoTable: Summary

In this section we provide a detailed analysis of the sensory streams provided by the VideoBench application and its deployment on the CoTable hardware infrastructure. This analysis provides us with a detailed tabulation of the technology characteristics of all of the sensory streams provided by VideoBench and CoTable. This bottom-up analysis, combined with our top-down analysis of the video editing task (Section 5.2.1), provide us with a detailed analysis of both the task and the technology domains of collaborative video editing. We explore the intersection of these two domains, through a simple case study below. A discussion of the utility of the sensory streams provided by VideoBench and CoTable in terms of meeting the communication needs of the video editing collaboration task described above is provided in Section 5.4.

5.3 The Case Study

In order to study how well the implementation of VideoBench meets the collaboration needs of the video editing task, we undertook a small case study. We analyze our experiences with using VideoBench (deployed on the CoTable hardware) within the HCI lab at the New Media Innovation Centre in Vancouver. This includes our experiences while developing and debugging the system, while demonstrating the system, as well as while carrying out structured video editing tasks. The structured video editing tasks were performed as part of the development process of the system (as a task to drive debugging the system) as well as for pre-study testing for the HCI gesture analysis carried out in [CFM+03].

Two colleagues from the HCI lab at the New Media Innovation Centre collaborated with the author to perform the video editing task described in Section 5.2.1. Informal discussions were held with the two users, exploring their experiences using the VideoBench and CoTable technology. Those experiences, combined with the author's experiences during the development, testing, and demonstrations of the system, make up the bulk of this analysis.

This case study is clearly ad-hoc and exploratory in nature. The goal was to gain experiences with two key aspects of the technology – gesture based interaction on tabletop devices and distributed artifact centric collaboration. In the following sections we briefly discuss our experiences with testing the VideoBench software on the CoTable

hardware. In Section 5.4, we consider our broader experiences with these technologies in the context of our analysis of the task and technology levels of the CoGScience Framework.

5.3.1 The video editing task

The task we used to test the CoTable and VideoBench systems was the simple video-editing task described in Section 5.2.1. Both of the users who assisted in testing the application were asked to perform the video editing task in each of three different configurations. The author was the second collaborator in each situation. The three configurations were:

- Co-located on the DT table – both users were collocated and performed the task using the DT table.
- Distributed with the user on the DT and the author on the desktop – the user performed the task on the DT while the author was working on a desktop computer using a mouse interface to the application.
- Distributed with the test user on the desktop and the author on the DT – the user performed the task on the desktop computer while the author was working on the DT.

Each user was taught how to use the VideoBench application before performing the editing task. The collocated task was then described and performed with the author acting as the collaborator. The component operations of the task were not assigned to individual users, and therefore the users had to communicate how they were going to perform the task. The author tried to let the user take the lead, but occasionally suggestions were made to remind the collaborator of the next step. After accomplishing the first scenario successfully, the user was introduced to the distributed environment, including the video feeds available on the various devices. The task was then repeated with the user on the DT and the author using the desktop environment. The user was then instructed on how to use the desktop application and the task was repeated a third time with the user on the desktop and the author using the DT.

The test environment for the user using the DT table is as described in Section 5.1.1 and shown in Figure 10 and Figure 12. The test environment for the desktop user is described in Section 5.1.2 and can be seen in Figure 11.

5.3.2 The Collocated Tabletop Experience

The collocated task was considered an effective collaboration environment. Users were able to talk to each other, they could see what each other were doing, instruction on what was desired was clear (through pointing, or duplicating gestures and movements), and turn taking was relatively straightforward. Subtle consequential communication, such as being able to see the user as they leaned forward to see something of interest, was easily visible.

As an experienced user, the author benefited from being able to see if the other users were struggling and therefore was able to give hints or guidance as to the next step in the task. It was much more difficult to determine if users were struggling in the distributed tasks. All of these observations speak to the value of the verbal, gestural, and body language communication channels that are inherent in the collocated tabletop environment. It is worth noting that these are both explicit (intentional) and consequential (unintentional) channels. Users indicated that seeing the other user's face during the task was of little importance, although they reported that gesture and body language were of significant importance.

Users noted a number of negative aspects of the collocated tabletop environment:

- One always had to be in contact with the DT contact pads (the user could not move around the table).
- One cannot hide his/her mistakes. This indicates that distributed interaction may have some benefits over collocated interaction, as suggested in [HL91, DES+00]. When collocated, everything is "on the table" so to speak.
- It felt like one person had to be in charge.
- Memorization of the techniques and gestures was necessary (and presumably non-trivial). Note that this is true in all configurations.
- There were some technical limitations in terms of consistency of gestural interaction on the DT table that caused frustration on the tabletop.

5.3.3 The distributed desktop experience

Users found tasks easy to perform on the desktop configuration, primarily because of the accuracy and robustness of the mouse interface over the touch interface. This seemed to make the actions easier to perform. All users reported that the **workspace visual stream** (the video feed of the tabletop, which provides an integrated task and personal space, or reference space) was critically important in their understanding of what the other user was doing. One of the users initially had the workspace visual stream quite small on the screen, and reported that once it was made larger it was more effective as a visual stream. This implied a trade off of screen real estate, with an increase in screen real estate for the low fidelity workspace visual stream and a decrease in screen real estate for the high fidelity **application visual stream** (the shared application, which provides only the task space) to the. One of the users reported that when he was not performing an action he was at least as likely to look at the workspace visual stream as he was to look at the application visual stream to see what the other user was doing. Users reported that the **personal visual stream** of the other user's head and shoulders (which provides a view of personal space) was not useful. Criticisms of the desktop environment focused on the size of the visual streams on the screen, the robustness of the software during synchronous actions, and the lack of utility of the personal visual stream of the tabletop user.

5.3.4 The distributed tabletop experience

The test case in which the user was using the tabletop and the author was using the desktop received the most criticism. Some of the positive aspects of the environment were the audio communication, the ability to see what was going on through the applications visual stream (gesture traces) and the ability to use the application visual stream to point to artifacts and clarify aspects of the communication. This last point is perhaps the most interesting.

In the design phase of the system it was decided not to provide the workspace visual stream from the desktop user to the DT user. Capturing video from the screen of the desktop user only provides an image of where the mouse pointer is located, while capturing video of the user's interactions with the mouse did not seem fruitful. This capability was missed. Desktop users essentially replaced the workspace visual stream with a gestural channel using the application visual stream. Rather than naturally pointing

at objects and indicating what was required with their hands (move this here, split this video clip), objects were selected or positions were indicated using the gestural traces in the application. One of the most fascinating things about this result is that this is a learned response. The users were adapting to the technological limitation of the system and finding a way to provide a gestural communication channel when it was not naturally supported by the existing visual streams. The rapidity with which this adaptation happened indicates that such a stream is likely to be important for artifact-centric collaboration.

Users reported the personal visual stream (head and shoulders) was rarely useful, as the users primarily used the audio and the application visual stream to coordinate actions and communicate. The only time the personal visual stream was reported as being used effectively was when the desktop user sat forward and put his chin in his hands, indicating very clearly that the user was waiting for the tabletop user to do something.

5.4 Discussion

In this section we consider the experiences described in Section 5.3 in the context of the CoGScience analysis presented in Section 5.2. The CoGScience Framework plays a critical role in mapping task **processes** and **channels** (the **task domain**) to **sensory streams** (the **technology domain**). Since our CoGScience task analysis indicates that many processes and channels were necessary for this task, we focus on expanding our bottom-up analysis of the sensory streams performed in 5.2.2. By considering the four sensory streams (one aural and three visual), we are able to provide a focused analysis of the system. A summary of the **technology domain** (**sensory streams** and stream types) and **task domain** (**channels** and **processes**) elements are given in Table 3. We provide a detailed analysis of these components below.

Technology Domain		Task Domain	
Stream Type	Stream	Channel	Process
Aural	Verbal	Verbal Paralinguistic	Verbalize (conversation) Turn-taking (conversation) Monitoring (conversation) Coordinate action (conversation)
Visual	Personal - head and shoulders	Facial expression Eye contact Body language	Awareness (conversation) Turn-taking (conversation, artifact) Monitoring (conversation, artifact) Intention (conversation, artifact) Coordinate action (artifact)
	Application	Workspace awareness Gesture (manipulation) Gesture (pointing)	Awareness (artifact) Turn-taking (artifact) Coordinate action (artifact) Manipulate (artifact) Modify (artifact) Intention (artifact) Monitor (artifact) Assistance (artifact)
	Workspace	Workspace awareness Body language Gesture (manipulation) Gesture (pointing) Gesture (kinetic) Gesture (spatial)	Awareness (conversation, artifact) Turn-taking (conversation, artifact) Coordinate action (artifact) Intention (artifact) Monitor (artifact) Assistance (artifact)

Table 3: Technology and Task domains – VideoBench on CoTable

5.4.1 The aural sensory stream

Our analysis of the technology characteristics of the **aural stream** (see Section 5.2.2.1) demonstrates that the stream is **high fidelity, full duplex, low latency**, and has **no echo**. In the video editing task, we consider the aural stream to support the **verbalization, turn-taking, coordinating action**, and **monitoring** processes (see Table 3). The **verbalize process** requires the **aural stream** to support both the **conversation** and the **work object (artifacts)**. That is, users carry out artifact-centric conversation as well as general conversation. Our analysis of the case study presented in Section 5.3 suggests that our users had few issues with using the aural stream for verbalization.

The **turn-taking** process uses the aural stream to effectively carry out a two way conversation. The **coordinating action** process uses the aural stream to effectively coordinate who is doing what during the video editing task. **Turn-taking** and

coordination were considered a problem by the users of the system, but this problem was indicated in both the collocated and distributed environments. Given that this problem occurred in the collocated condition, we have no indication from the users in our case study that the use of the aural stream in the distributed condition causes a problem in performing these processes.

The **monitoring** process also requires the aural stream. Interestingly, the aural stream is primarily used for monitoring through its absence, with the monitoring function prompting users to act when there were "awkward silences". Note that this use does not appear to indicate a problem with the aural stream itself, but instead seems to imply a lack of ability to monitor using the visual streams.

5.4.2 The personal visual sensory stream

The goal of the personal **visual sensory stream** was to create a personal space that replicates the "across the table" feel of the face-to-face tabletop environment. We classify the personal **visual stream** as being an important mechanism for providing the **facial expression**, **eye contact**, and **body language** communication **channels** (see Table 3). These channels are important in enabling the **awareness**, **turn-taking**, **monitoring**, **coordinating,** and **communicating intention** processes. From a technical perspective, the personal visual stream does a moderate job at communicating **facial expression**, although subtle facial expressions may be difficult to discern because of the low **fidelity** of the visual stream. Similarly, **body language** is communicated relatively poorly, as although it may be possible to recognize when the remote user reaches for or points at an artifact, these cues are not strong. A basic level of **eye-contact** is provided, with the location of the cameras close enough to the screen so that when a collaborator looks at the LCD screen, he/she is looking close to the camera.

Users almost exclusively implied that the personal visual stream was not an important sensory stream for the task. The main use that users reported for this visual stream was **coordinating action** when no actions were being performed on the table (an "awkward pause" as defined by one user). In this instance, the user might glance at the personal visual stream to discern if the collaborator was about to perform an action.

5.4.3 The application and workspace visual sensory stream

Both the application and the workspace **visual sensory streams** provide information about task and reference space. We therefore consider them together, allowing us to contrast and compare how they provide the human **communication channels** required for the video editing task. We classify the application visual stream and the workspace visual stream as being important mechanisms for providing the **workspace awareness**, artifact **manipulation gestures**, and artifact **pointing gestures**. In addition, we classify the workspace visual stream as also providing **body language**, **kinetic gestures** (hand waving), and **spatial gestures** (showing relationships) (see Table 3). It is important to point out that both visual streams provide **explicit communication** (communication done intentionally) while only the workspace visual stream shows **consequential communication** (unintentional actions). That is, any action communicated by the application visual stream is the result of an explicit, intentional action such as touching the tabletop or clicking the mouse. The workspace visual stream, on the other hand, shows the user's hands and body if they are placed over top of the table, and therefore can be used to show both explicit and consequential communication.

Both visual streams are utilized to support a range of other task processes, including **awareness**, **turn-taking**, **coordinating action**, **communicating intention**, **monitoring**, and providing **assistance** (see Table 3). Note that these processes are primarily related to the **work object** rather than the **conversation,** and are useful at providing **workspace awareness**. The application stream also supports the **manipulating** process (artifacts). Since the workspace visual stream also provides **kinetic gesture** and **spatial gesture**, it supports processes that relate to the **conversation**. That is, one of the primary uses of kinetic and spatial gesture is to **coordinate speech**, and therefore the workspace visual stream is the only mechanism available in VideoBench to support such a communication process.

Further evidence of the importance of the work space, and gestural communication in particular, to this task is demonstrated by the fact that users created a **gestural communication channel** when it was not provided as part of the environment. For example, when it is necessary to **assist** a user (a communication **process**), it is often necessary to **point** at an artifact (a communication **channel**). When the desktop user

needs to assist the tabletop user, the system does not provide a visual stream that captures pointing gestures made with the hand (there is no workspace visual stream). Thus in order to communicate **assistance** effectively, it is necessary to create a **gestural pointing channel** using another **visual stream**. Users created a **pointing gesture channel** using the application **visual stream**, utilizing the visual traces to circle or underline an artifact of interest and/or selecting an artifact, as part of the **process of assisting** the user.

The importance of **gestural channels** is further demonstrated when considering how users used the workspace **visual stream**. The initial desktop configuration had the high **fidelity** (resolution) application **visual stream** (1024x768 pixels) filling the entire screen with the personal **visual stream** and the workspace **visual streams** visible as small windows in the corners (352x288 pixels, as seen in Figure 11). On several occasions users preferred to enlarge the workspace **visual stream** despite its low **fidelity**. Note that enlarging the video window increases the space the visual stream takes up on the screen, but does not increase the **fidelity** (it is still only 352x288 pixels). This implies that the information that is communicated by the workspace **visual stream** is important enough that users are willing to accept a decrease in the **fidelity** (by choosing the low resolution stream over the high resolution stream) in order to ensure that they receive that information. It also implies that the utility of the workspace **visual stream** may be a result of its ability to capture a broader range of communication channels (workspace awareness, some body language, and the full range of gesture channels).

5.5 Summary

The goal of this research was to explore the issues and problems that arise when a natural interaction environment, such as that presented to users when communicating over a tabletop is impacted by having one or more of the users communicating from a remote location. We therefore created CoTable, a distributed tabletop collaboration environment that provided us with a technology environment to explore distributed, artifact-centric collaboration. We gained experience with this environment by carrying out a case study of a distributed, video editing application (VideoBench) implemented on the CoTable system. We chose the VideoBench application for this research because it embodies an artifact-centric collaboration task, it is visual in nature, and the application directly supports gestural interaction. Collaborators need to be able to communicate

effectively about the artifacts being manipulated, both in order to coordinate the collaboration task as well as to communicate information about the artifacts.

One of the most significant contributions of this chapter is the application of the CoGScience Framework to VideoBench and CoTable. In particular we performed:

- A top-down analysis of the **task domain** of the video editing task for both collocated and distributed collaboration (Section 5.2.1);
- A bottom-up analysis of **technology domain** of the VideoBench software running in the CoTable hardware environment (Section 5.2.2); and
- The **intersection** of the **task domain** and **technology domain** in the context of our experiences with using VideoBench and CoTable (Section 5.4).

Our analysis of collocated video editing on a tabletop device reveals the following:

- The video editing task chosen consists of a set of sub-tasks and that those sub-tasks span a number of **task types** (**creative, decision,** and **perform**);
- Accomplishing the task successfully requires switching sub-tasks effectively;
- The task types span several CoGScience **task quadrants**;
- The sub-tasks require many **processes**; and
- The task processes require many (almost all) communication **channels**.

In general, our application of CoGScience to the task domain of video editing suggests that artifact-centric collaboration, because of its richness in communication, will often require many (if not all) human communication channels listed in the CoGScience Framework. Thus, considering an artifact-centric collaboration task within the CoGScience Framework is not a matter of eliminating channels that do not apply, but instead prioritizing channels depending on the needs of the task.

Our analysis of the intersection of the task domain with the technology domain, in the context of our case study, suggests the following about distributed, artifact-centric, video editing:

- The **verbal stream** provided both **verbal** and **paralinguistic channels** and adequately supported the **verbalization**, **turn-taking**, **monitoring**, and **coordinating processes** required by the task.
- The **personal visual stream** provided moderate **facial expression**, **basic eye contact**, and **poor body language channels**. The personal visual stream was not

used very much, suggesting that it did not contribute significantly to the **awareness**, **turn-taking**, **monitoring**, **intention**, and **coordinating processes** required by the task.

- The **application visual stream** provided a moderate workspace **awareness** and **manipulative gesture channel**. It provides a very basic **pointing gesture channel**. Our experiences suggest that although the application visual stream was used extensively for the **manipulate** and **modify** processes, it appears that it was only partially used for **awareness**, **turn-taking**, **coordinating**, **monitoring**, and **assistance**. Our analysis suggests that this may be because the application visual stream does not communicate a wide range of communication channels. Our experiences infer that the lack of **body language**, **kinetic gesture**, **spatial gesture** (and in general any **consequential communication**) reduced the effectiveness of this visual stream.
- The **workspace visual stream** provided moderate workspace **awareness**, **body language**, and **gesture channels**. We classify these channels as moderate because, although they communicate the necessary information, they do so at a low **fidelity** (resolution) and **quality** (low quality due to compression). Our analysis suggests that the **workspace visual stream** was important in supporting the **awareness**, **turn-taking**, **coordinating**, **intention**, **monitoring**, and **assistance** task **processes** (primarily related to the **artifact**, not the **conversation**).

Perhaps the most important results from this work are the indications of how important gestural communication channels are for artifact-centric collaboration. Although our case study suggests all types of gesture are important, gestures relevant to artifacts are of particular importance. In particular, we found that:

- **Visual streams** that provide **gesture** are used often.
- When gestural communication **channels** are not provided, users find ways to communicate the required **gestures** using other **visual channels**.
- Users appear to prefer **low fidelity visual streams** that encode multiple **gestural**, **body language**, and **awareness** communication channels over **high fidelity** streams that encode fewer communication channels. This suggests that the richness of the visual streams may have more value than the fidelity of the

information communicated when concerned with artifact-centric, gestural interaction.

Despite the interesting issues raised in this Chapter, it is important to discuss the limitations of this case study. We discuss our experiences with CoTable and VideoBench, but the investigations are not a rigorous study. Our analysis is participatory in nature and based on the experiences gained while developing, testing, and demonstrating the CoTable hardware and VideoBench software. Out of necessity, this development and testing required multiple people, and it is the interactions with these colleagues in the lab that have provided us with our analysis. The author of this dissertation was a key participant in the interactions users had with the CoTable system, introducing a potential bias in the interactions that users had with the system. The author prompted and directed the collaborating users, directing them through the task being carried out. Because this study was a participatory activity, it is important to take this into consideration when assessing the results presented above.

Although providing us with interesting insights into the use of these technologies, as with many qualitative studies, this research has resulted in the creation of more questions than it has provided answers. This research was very effective in one regard. It made it quite clear that artifact-centric collaboration is highly complex, and there are a myriad of subtle interactions that need to be considered. Indeed, it is our need to understand these interactions that inspired the creation of the CoGScience Framework. Ultimately, our interests lie in distributed, artifact-centric scientific collaboration, and in many ways CoTable provides much of the inspiration for the remainder of the research presented in this dissertation (see Section 1.1.1 for "the story"). In fact, the foundations of many of our objectives and research questions (Section 1.2), some of which are listed below, are based on questions that arise from the research presented in this chapter.

Objective 2: Develop a deep understanding of how scientific researchers interact with digital artifacts when they collaborate.

- *Question 1: What role do digital artifacts play in scientific collaboration?*
- *Question 2: What information is lost when such collaboration takes place at a distance?*

- *Question 3: What communication channels are used to encode information during artifact-centric collaboration?*

Objective 3: Evaluate advanced collaboration modalities and technologies for scientific collaboration.

- *Question 1: How do researchers use advanced collaboration technologies?*
- *Question 2: How well do those technologies work?*

Objective 4: Develop a set of design guidelines for the development of effective collaboration tools for scientific researchers.

- *Question 1: What human communication channels need to be supported for artifact-centric collaboration?*

The remainder of this dissertation attempts to provide new knowledge that helps to answer these questions.

6 Scientific Collaboratories in Action – An Analysis

In Section 2.1 we discuss the importance of collaboration in the computational sciences. We also highlight the fact that detailed analyses of the use of collaboration technologies in the sciences are relatively rare. Media Spaces [Har09, Section 2.3.3.3], as both a technology and a social facilitator, have the potential to meet many of these needs. In this chapter, we focus on the use of Scientific Media Spaces (SMS) as a tool for supporting collaboration in scientific research. In particular, we discuss the design, deployment, and use of a set of SMS environments deployed by the Centre for Interdisciplinary Research in the Mathematical and Computational Sciences (IRMACS) from January 2005 to December 2009. This Chapter attempts to contribute knowledge to two of the four research objectives listed in Section 1.2.

- *Objective 1:* Develop a broad understanding of how scientific researchers collaborate.
- *Objective 3:* Evaluate advanced collaboration modalities and technologies for scientific collaboration.

In particular, we consider the following research questions:

- *How do collaboration patterns change in the presence of technology?*
- *How do researchers use advanced collaboration technologies?*
- *How well do those technologies work?*

This chapter proceeds as follows:

- Section 6.1 discusses media spaces and their applicability in supporting the computational sciences.
- Section 6.2 describes WestGrid and IRMACS, the two collaboratories studied in this Chapter.
- Section 6.3 describes our quantitative analysis of the IRMACS SMS usage from 2005 to 2009.
- Section 6.4 discusses the anecdotal experience we gained in the design, implementation, and operation of the IRMACS SMS infrastructure from 2003 to 2009.
- Section 6.5 summarizes the findings from this research.

This chapter is an extension of the analysis presented in "Build it – will they come: Media Spaces in the support of Computational Science" [CZ09], a chapter from the book *Media Space: 20+ Years of Mediated Life* [Har09].

6.1 Scientific Media Spaces

In the remainder of this chapter, we describe our experiences in planning, building, and operating an extensive SMS infrastructure in support of the computational sciences in Western Canada. Recall that Media Spaces facilitate the creation of *place* from *space* (see Section 2.3.3.3 for details on Media Spaces). We use the media space concept to encapsulate the process of creating a *place* for scientific collaboration from a set of technologies that exist in one or more physical *spaces*. Below, we focus on the concept of creating a Scientific Media Space from the set of collaboration technologies that exist in the technologically sophisticated collaboration rooms that were created at the IRMACS Centre.

6.1.1 Media Spaces in the Sciences

Communication is a fundamental component of supporting distributed collaborative science. If the goal is to create a large, distributed laboratory, a question follows naturally; can media space technologies be leveraged to provide an environment that meets the collaboration needs of a distributed scientific community. In many ways, early media space research environments were Scientific Media Spaces. The PARC [HBA+97] and EuroPARC [BD97] media space systems were created to support scientific research in computer science. They brought together distributed communities of researchers working on a common project. This community was distinct in that the research area being studied was in fact the media space environment itself. The question we explore in the remainder of this chapter is in what ways a media space environment might support the general scientific community.

Figure 15: A theatre (left) and meeting room (right) Scientific Media Space

6.1.2 AccessGrid as a Scientific Media Space

AccessGrid (AG) is a technology platform for the support of distributed, scientific collaboration [COP+00]. The original vision of AccessGrid was in many ways as a modern scientific media space. Designed to support group-to-group collaboration in a room-based environment, AccessGrid directly supports multiple, high resolution, large screen displays, multiple cameras, high quality acoustically echo cancelled full duplex audio, and a range of interaction technologies. The ideal collaborative environment envisioned by the AccessGrid developers consisted of *"an intentionally designed space, one that would be rewarding to be in, one that provides a sense of copresence with other groups using similar spaces. We envision a space with ambient video and audio, large-scale displays and with software to enable the relatively transparent sharing of ideas, thoughts, experiments, applications and conversation. We envision a space where we can 'hangout' comfortably with colleagues at other places, and then use the same space to attend and participate in structured meetings such as site visits, remote conferences, tutorials, lectures, etc."* [COP+00]. A more detailed discussion of AccessGrid and how it can be used in SMS environments is given in [CZ09].

6.2 Collaboratories in Western Canada

Our analysis in this chapter focuses on two collaboratories in Western Canada, WestGrid and the IRMACS Centre. We provide a brief description of these two collaboratories below.

6.2.1 What is WestGrid?

WestGrid is a large computational science consortium in Western Canada, spanning the four westernmost Canadian provinces or roughly half the country geographically.

According to collaboratories classification of Bos *et al.* [BZO+08], it is a Community Infrastructure Project, providing a set of infrastructure (software tools, protocols, instruments, computers) that facilitates science. WestGrid provides computational science resources (high performance computers, data storage, networking, collaboration, and visualization technologies) to over 1000 researchers. An important aspect of the computational science infrastructure that WestGrid has created is a set of SMS environments. These media spaces are designed to provide distant collaborating researchers with the ability to communicate effectively with colleagues across campus, across the country, and around the world. Although we do not analyze the use of the WestGrid infrastructure in detail, it is introduced here because many of the remote collaborations that are held within the IRMACS SMS environments connect to facilities at other WestGrid institutions.

6.2.2 What is IRMACS?

At approximately the same time as WestGrid was deploying its SMS infrastructure, another related research centre was in the implementation stages. Funded in 2002 and becoming operational in late 2004, the IRMACS Center at Simon Fraser University (SFU) has taken a somewhat novel approach to supporting interdisciplinary research. The IRMACS Centre supports research across a wide range of disciplines by creating a physical "meeting place" for its research community (creating 25000 sq. ft. of open office, lab, and meeting space) as well as providing the technological infrastructure to perform that research in as effective a manner as possible. In this sense, it can be thought of as a Community Infrastructure Project [BZO+08].

One of the key goals of the IRMACS Centre is to break down the distance barrier that is so often a problem in collaborative research [OO00]. IRMACS is a *space* specifically designed to promote and foster collaboration in the computational sciences. It provides an attractive environment, both architecturally and socially, drawing researchers to the Centre through its physical space (labs, meeting rooms, and presentation studio), its technology (computation, storage, meeting rooms, collaboration, and visualization), and perhaps most importantly, its research community. IRMACS extends its physical space to remote interdisciplinary collaborators using SMS environments. From the perspective of media space research, the goal of IRMACS is not to create a **space** for interdisciplinary

research, but instead to create a **place** that draws interdisciplinary researchers together academically and socially, both locally and at a distance.

6.2.3 IRMACS Scientific Media Space Design

We targeted the IRMACS SMS infrastructure at meeting the collaboration needs of a wide range of scientific users. From an SMS perspective, this presents an interesting design problem. Most media space environments are targeted at a single community need, and therefore can be customized to support a community of practice. The IRMACS infrastructure must address user needs across a wide range of scientific communities and across a wide range of collaboration scenarios. Although one of the defining properties of a media space is its ability to support a range of needs, the IRMACS diversity of use amplifies this requirement. The SMS infrastructure needs to be highly configurable and at the same time maintain simplicity of use.

Distance collaboration was always a fundamental part of the IRMACS vision. Leveraging the fact that IRMACS was one year behind WestGrid in its funding and implementation of its SMS infrastructure, IRMACS was able to provide one of the most advanced SMS environments of the WestGrid institutes. The IRMACS SMS rooms are designed meeting room spaces (they were designed as part of the building construction and are not retrofitted meeting rooms), with the ability to be used as both traditional collocated or distributed collaboration spaces. It was recognized early on in the IRMACS design that we did not want to build spaces for distance collaboration, but instead spaces that were designed for collocated collaboration that could be utilized for collaborating at a distance at the "touch of a button". That is, we wanted to have the ability to transform a designed collaboration space (e.g. a meeting room) to a scientific media space on demand.

IRMACS created six SMS rooms, with rooms designed to fill specific roles in terms of the type of collaboration that they support. Rooms range from a 75-seat lecture theatre (with a high-resolution stereoscopic 3D visualization capability), through traditional meeting rooms, to a lab scale shared scientific visualization laboratory (see Figure 15). Each IRMACS SMS has high quality acoustic echo cancellation, multiple displays, and multiple cameras. Like the rest of the physical environment at IRMACS, the SMS rooms are designed to be both physically appealing as well as technologically sophisticated. The

rooms supported a wide range of collaboration technologies, including traditional teleconferencing, video conferencing (H323), and desktop collaboration technologies such as AccessGrid, iChat, Skype, and VNC. In addition, all of the IRMACS SMS environments make use of touch sensitive screen overlays (Smartboards), allowing users to interact with applications by directly touching the screen or annotating documents by writing on the screen with a digital pen.

The IRMACS SMS infrastructure has been consistently upgraded since IRMACS first came on line in late 2004, with new technologies such as high definition H323 video conferencing (HD 1080p cameras, encoding, transmission, and display), full HD displays (1920x1080 pixels), and room control systems (LCD panels that control the technology) being deployed in several of the IRMACS SMS rooms. Much of this new technology has been driven by the need to make the rooms more usable. As we learn about how users utilize the technologies in the rooms, we have been able to customize the infrastructure such that particular collaboration scenarios can be realized at the touch of a button.

In the remainder of this chapter, we explore the community of users that have taken advantage of the IRMACS SMS infrastructure over the past five years (2005 – 2009). We discuss our observations of how the IRMACS SMS rooms were used, the impacts that usage had on design changes in the SMS rooms over time, and our observations of the impact these changes have had on the collaboration patterns of IRMACS researchers since the Centre became operational. Note that this is a high-level analysis of a wide range of scientific collaborations. For an in-depth study of how a single scientific research group makes use of advanced collaboration technologies, see Chapter 7 and related papers [CS05][CS07].

6.3 Analysis of SMS in Action: We Built It – Did They Come?

The IRMACS SMS infrastructure was created because IRMACS recognized that scientific researchers had a need to collaborate with remote colleagues and that this was an important aspect of the emerging computational science research community. This need was not as well defined as perhaps it could have been, and the infrastructure was in some sense created with a “build it and they will come” approach. We recognized a need, but did not understand the usage pattern of this community well. We therefore created a

flexible and extensible infrastructure, while at the same time attempted to hide the complexity of the collaboration infrastructure.

6.3.1 What is Distributed Collaboration?

Before analyzing usage patterns and our experiences with operating a collaboration infrastructure on the scale of IRMACS, it is important to discuss how we define a remote collaboration. Recall that the IRMACS SMS rooms are highly sophisticated, supporting a range of collaboration technologies. So what constitutes a remote collaboration? Considering this from a CoGScience Framework perspective, we utilize the task domain, the technology domain, and group distribution characteristic aspects of the framework to define remote collaboration. We define a remote collaboration as any activity or **task** that utilizes one of the IRMACS SMS rooms (and therefore the **technologies** in that room) to bring collaborators at two or more **distributed** sites together to accomplish that task.

Like task and technology, the **distribution** of users varies dramatically across our SMS users, ranging from two people at two sites to over 130 people at 22 sites. Any distributed collaboration scenario between these two extremes fits our definition of distributed collaboration and is therefore considered as part of the analysis carried out in this chapter.

We do not restrict the activity or **task**, but consider any activity that uses collaboration technologies in one of the SMS rooms as relevant to this analysis. This can range from a simple teleconference between two sites to a remote seminar involving tens of remote institutions and over 100 people. Clearly, the latter is an example of an extensive remote collaboration, but one might question the validity of calling a teleconference an SMS meeting.

We include teleconferencing for two reasons. First, a high **fidelity**, **echo** free **aural stream** is critical to almost all collaboration tasks. In our experience, if a telephone capability is not included in an SMS room, it limits the range of collaborations the room can support. This is particularly important in “problem” collaboration scenarios, such as when a remote collaborator who is travelling (e.g. at a conference) needs to join a research meeting. Second, although many of our distributed collaborations rely on the telephone to provide the aural stream, they also utilize other visual sensory streams. These might include video streams of remote collaborators, a visual stream of a shared presentation, or a visual stream of a shared interactive Smartboard session. Thus,

although such an SMS session may "only" be using the telephone, the collaboration itself can be fairly sophisticated (see Chapter 7 for examples).

6.3.2 Data Extraction and Analysis

All information about meeting room use, for both collocated and distributed meetings, is extracted from the booking calendars of the IRMACS Centre. Fortunately, since its inception in 2004, IRMACS has been focussed on remote collaboration, and therefore its calendar system tracks the use of remote collaboration technologies. Such tracking is necessary for grant reporting and grant renewal, and therefore has remained high on the priority list throughout the project. When a room is booked, the technical requirements of the booking are noted, including whether the meeting requires tele-conferencing, video conferencing (H323, AccessGrid, Skype, iChat etc.), projectors, Smartboards, and other relevant technology. The type of meeting (research meeting, seminar, PhD defence, etc.) is also recorded. This makes it relatively easy to extract the number of remote collaborations that have been held in IRMACS since 2005.

It is important to stress that data extraction has been *relatively* easy! Like any database system, the IRMACS calendar depends on the quality of the data entered into the system. Problems in this regard include data not being entered correctly, the data that is provided for entry not being correct, and not being provided with data at all. This is exacerbated by the success of the technology and the growing expertise of our users. For example, research groups regularly use IRMACS meetings rooms for their project meetings (see Section 6.3.5 for statistics). If for some reason one of the group members cannot make it to the meeting (they are at home, on another campus, or at a conference), they can be connected to the meeting at a push of a button (or two). Since some IRMACS users have been doing this for five years, the use of the SMS technologies has become almost trivial. Users no longer need technical help, and many use the collaboration technologies as seamlessly as they do their computers. Such a meeting would never get entered into the booking system, and therefore the IRMACS meeting room calendar underestimates the number of distributed booking in this regard.

The opposite is also true. Some of our research groups book recurring meetings (weekly, monthly) and regularly use the distributed SMS technologies. These are diligently entered into the booking system. Unfortunately, not all meetings that are

booked as distributed meetings are actually distributed. For example, a researcher that normally attends a research meeting from the University of British Columbia (also located in Vancouver) might sometimes be visiting SFU and attend a meeting physically. Or a remote collaborator at the University of Saskatchewan who usually attends a remote meeting might be busy during a given meeting and choose not to attend remotely. In both cases, the meeting will erroneously be recorded as a distributed SMS meeting. Our meeting room bookings therefore over-represent the true number of distributed collaborations to some degree.

We manage these issues in two ways. First, researchers are reminded to accurately book their rooms in terms of the technology they require. When we notice that research groups are booking rooms and know they are using collaboration technology, we often follow up with them to ensure that bookings are entered correctly. Second, for those bookings that we believe are overbooked, we directly contact the project leader and ask them to provide feedback on the number of bookings that actually use distributed technology. Almost all research projects provide us with feedback. Typical feedback consists of responses that allow us to adjust frequency (e.g. one in every three meeting were remote), adjust time spans (e.g. we used it every meeting from January to March but not in April), or correct/confirm the booking (e.g. confirm that either all or none of the meetings were distributed). It is clear that we cannot be 100% confident in the accuracy of our booking statistics, but at the same time we have gone to significant effort to be as accurate as possible.

6.3.3 Who Uses IRMACS?

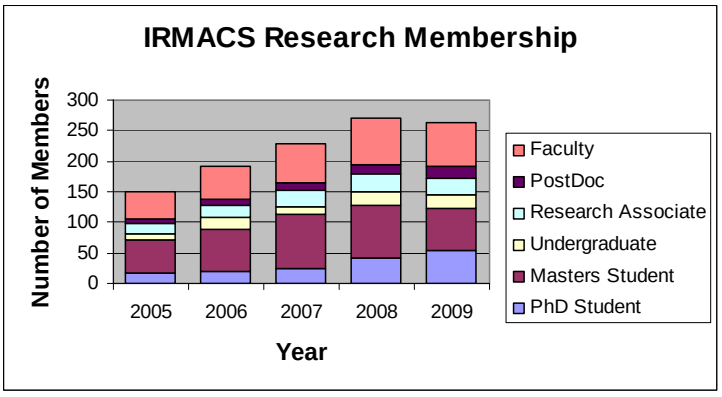


Figure 16: IRMACS Research Memberships 2005 - 2009

IRMACS membership is based around research projects. Faculty members apply for research projects within the Centre, and research projects sponsor individual research members. Over the 2009 calendar year, IRMACS hosted 58 research projects and 262 researchers, with research projects spanning 16 disciplines and departments. Since its inception in late 2004, IRMACS has hosted 81 projects and 570 researchers. Note that this does not include researchers from other institutions who visited IRMACS and used the facilities (28 in 2009), researchers who are involved in IRMACS projects from other institutions (24 in 2009), and conference visitors (in the hundreds). The number of research members, grouped by year horizontally and by the type of membership vertically, is given in Figure 16. The number of research projects from 2005 to 2009 is given in Figure 17.

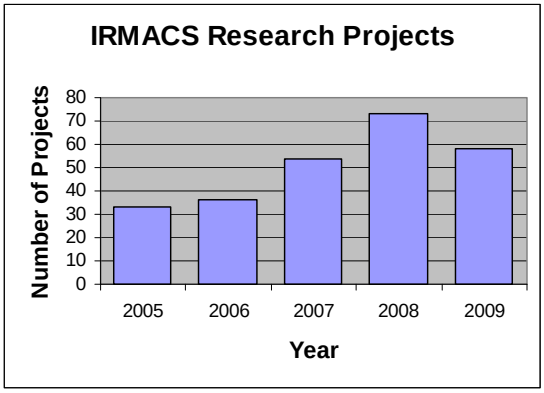


Figure 17: IRMACS Research Projects 2005 - 2009

It is worth pointing out that the research membership at IRMACS is quite dynamic. For example, even though there are a smaller number of research projects in 2009 than 2008 (58 and 73 projects respectively), the number of researchers who were members during the two years is roughly the same (262 and 270 researchers respectively). In addition, because of the number of projects and researchers involved in IRMACS, researchers are arriving and leaving IRMACS on regular basis. On any given day, IRMACS typically has between 200 and 250 active research memberships (see Figure 18). Over the span of a year, from 70 to 100 researchers will both join and leave IRMACS (bringing the total number of researchers involved in IRMACS in a given year to approximately 270 as described above). For example, in the 2009 calendar year, 75 researchers left IRMACS

while 101 researchers joined. This turnover is primarily graduate students, post doctoral researchers, and research associates.

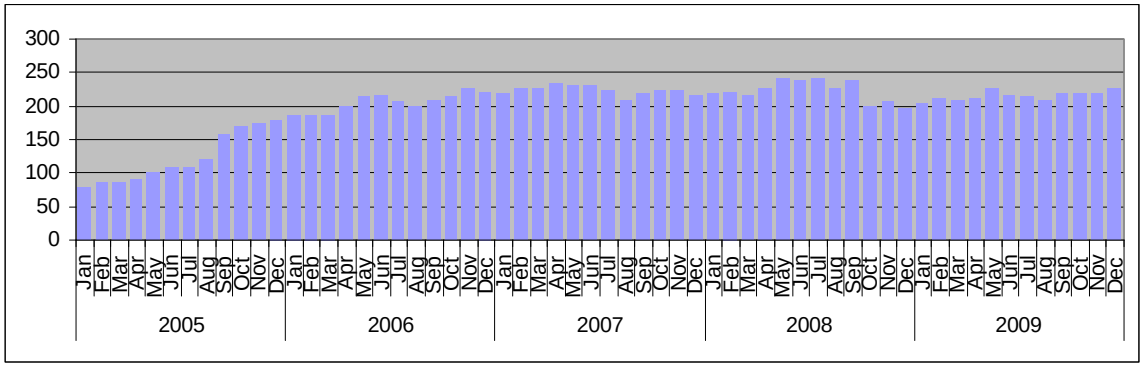


Figure 18: IRMACS Membership on a monthly basis.

The number of IRMACS researchers has reached a relatively constant level. As can be seen from the graph in Figure 18, the number of active IRMACS researchers has been hovering between 200 and 240 since March of 2006. Although the IRMACS annual research membership has been growing, this is attributed more to the increase in the number of researchers that come and go from the IRMACS Centre rather than the number of research memberships that are active at any one time. The limitation of the number of researchers is at least partially driven by the amount of IRMACS lab space. IRMACS researchers share 100 workstations (desks with computers) spread across four labs. The bulk of the researchers in IRMACS are graduate students, post doctoral researchers, or research assistants (see Figure 16), and many use the IRMACS labs as their university office (they often do not have other department offices). Many IRMACS projects also use the IRMACS facility (labs and meeting rooms) as their home base, rather than utilize individual departmental offices and meeting rooms.

Note that because of the use of SMS technologies, the number of collaborators (researchers at other institutions who collaborate with IRMACS research projects) that the Centre supports is relatively unbounded. Unfortunately, we have been unable to find a way to accurately gauge the number of collaborators whose research is facilitated through the use of IRMACS SMS rooms. As discussed in Section 6.3.5, IRMACS hosts a very large number of distributed, scientific collaborations. Exactly how many remote researchers this impacts is unclear, with the only fact we know for sure is that there is at least one person at the other end of each distributed collaboration.

6.3.4 What Do They Come For?

The usage pattern of the WestGrid and IRMACS SMS infrastructure is a complex one. The SMS infrastructure is used for a wide range of purposes, by a wide range of users, each with a wide range of experiences with using the technologies. Some of our research groups use the SMS rooms for distant collaborations several times a week, while other researchers use the rooms once (e.g. for a PhD defence) and never use them again. Some uses are for formal presentations to a large and widely distributed audience (left image in Figure 15) while others are informal, exploratory, and often intense research group meetings with only one or two distant collaborators. We can decompose our major SMS usage into four broad categories:

- *Research meetings*: One of the primary uses of scientific media spaces is to support scientific research meetings (over half or 254 of the IRMACS SMS meetings in 2009 where remote research meetings). IRMACS SMS environments are used to support a range of research groups that span WestGrid, Canada, and the world. In many cases, these collaborations involve the joining of two or more physical SMS environments while at other times the meeting may only involve one or two remote researchers joining a larger group of collaborators in a single SMS environment. These collaborations are usually interactive in nature and many of them revolve around the sharing of digital artifacts. We explore this type of collaboration further in Chapter 7.
- *Research dissemination*: In addition to supporting research meetings among individual research groups, SMS are also used to disseminate research results to a wider audience. For example, IRMACS leads (with Dalhousie University) the Coast-to-Coast (C2C) Seminar Series [BJL+06], a bi-weekly seminar series that brings together researchers from across Canada to present and discuss their research. These sessions are interactive presentations that involve up to 22 sites across the country, with upwards of 130 attendees at some sessions. This activity has been occurring since 2005. We perform a detailed analysis of remote presentations in Chapter 8 through Chapter 11.
- *Training meetings*: WestGrid technical support staff use SMS for providing advanced training courses to the scientific community. Courses are given on

using, programming, and optimizing computational algorithms for high-performance computing systems, using scientific visualization, and using collaboration technologies. These sessions typically involve many of the fourteen WestGrid institutes and can have anywhere from one to twenty participants at a given site. The sessions are typically interactive in nature, with the training sessions often involving live interactive demonstrations that are shared between all sites. The author was often involved in either delivering or supporting these sessions. Such training sessions have been delivered since 2004.

- *Operational meetings*: WestGrid, as a distributed consortium, uses the SMS infrastructure for operational purposes including financial, technical, and strategic meetings. These meetings involve all WestGrid institutes with one to four people at each site. They have been occurring on a regular basis since 2004.

6.3.5 How Often do They Come?

The vision of an SMS infrastructure that supports a wide range of scientific uses is ambitious in its scope, but over the six-year period of its planning, deployment, and use we view the IRMACS infrastructure deployment as fundamentally successful. There are many aspects of the SMS infrastructure that could be more effective, but the increased frequency of use of our users, the increasing number of users using the facilities, and the level of sophistication demonstrated by our users all indicate that the SMS infrastructure is increasingly meeting the needs of our users.

In 2004, WestGrid and IRMACS infrastructure significantly changed the collaboration landscape at SFU. One key dimension of this change was the target audience of the SMS infrastructure. Before the IRMACS Centre was established, the primary use of collaboration technologies at SFU was targeted at either remote teaching or administrative meetings. The IRMACS infrastructure is almost exclusively focused on supporting distributed scientific research. A second equally important dimension of change was the number and capability of the SMS spaces that were built by WestGrid and IRMACS. The main SFU campus went from having one room and one mobile video conferencing space to having an additional seven technologically sophisticated SMS rooms.

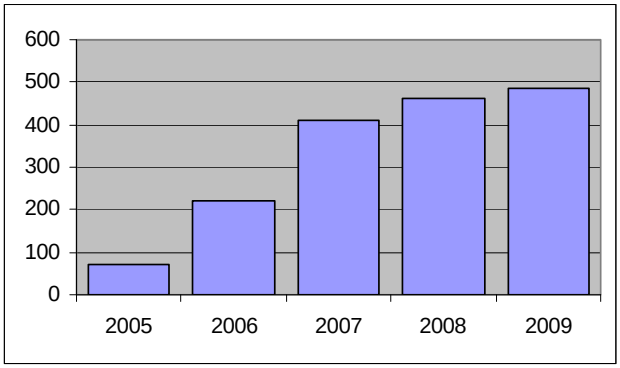


Figure 19: Number of IRMACS SMS Meeting 2005 - 2009

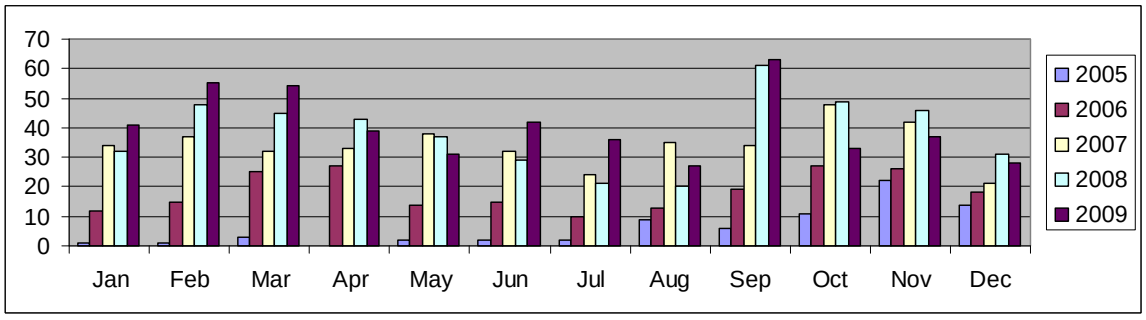


Figure 20: Number of monthly IRMACS SMS Meetings, broken down by year

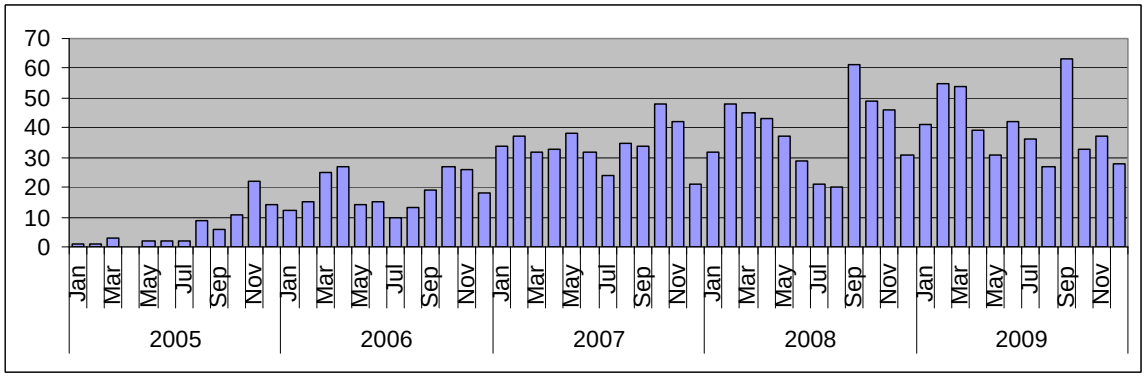


Figure 21: Number of yearly IRMACS SMS meetings, broken down by month

6.3.5.1 The IRMACS SMS Infrastructure

The use of the IRMACS SMS environments for distant collaboration has steadily increased since IRMACS opened in January of 2005 (See Figure 19). At the end of its first year of operation (the end of 2005), IRMACS supported 33 research projects and 150 researchers. During the 2005 calendar year, 73 distributed SMS meetings occurred

with a steady growth in the number of SMS meetings as the IRMACS community grew in size and as the researchers became familiar with the IRMACS SMS capabilities. By the end of the 2009 calendar year, the IRMACS research community grew to 58 projects and 262 researchers. The IRMACS SMS spaces were used for distant collaboration 486 times during 2009 for approximately 806 hours of SMS collaboration time (on average 9.35 meetings and 15.5 hours of SMS time per week). During the same calendar year, the IRMACS meeting rooms were used for a total of 1,819 meetings (both traditional and remote SMS sessions) and a total of over 3,615 hours of meeting time. This implies that approximately 27% of the meetings (and 22% of the number of meeting hours) in the IRMACS facility had remote collaborators participating through use of the IRMACS SMS infrastructure.

There are some interesting features in the above data that are worth considering in more detail. First, there is a clear trend that the amount of collaboration has been steadily increasing since 2005, although it appears that the speed of this increase is tapering (see Figure 19). For example, the 2007/2008 and 2008/2009 increases in SMS meetings are smaller than the 2005/2006 and 2006/2007 increases. These increases can also be observed in the SMS meeting increases on a per month basis. That is, in almost all months there is a clear increase in the number of SMS meetings when considering subsequent years (Figure 20).

If one considers the number of researchers associated with IRMACS on a yearly basis, a similar increase can be seen (comparing Figure 16 and Figure 19), and there is indeed a strong correlation between number of researchers and number of SMS meetings (R^2 = 0.964). It is therefore tempting to attribute increased SMS usage to the increasing user community. This would not be correct. Recall that from January 2005 (when IRMACS opened) to April 2006 there was a relatively steady monthly increase of both research membership and SMS usage (see Figure 18 and Figure 21). Indeed, there is a strong correlation between the two (R^2 = 0.748). In contrast, from April 2006 onwards, the monthly size of the IMRACS research community was relatively stable (mean = 219, σ = 11.2). Comparing the number of active researchers with the number of SMS meetings in each of the months from April 2006 to December 2009 shows almost no correlation

between the two ($R^2 = 0.005$). This analysis implies that the number of researchers, although possibly a contributing factor, is not the main driver of SMS usage growth.

Although our current data does not provide us with any empirical evidence, our observations suggest two other driving factors for the frequency of SMS meetings. The first is that the number of SMS meetings is driven by research activity. During the months in which research activity at the university tapers off (during June, July, August, and December) we see less SMS activities. During the first three months of both the fall and the spring terms (September to November and January to March) when research activity increases (graduate students start courses, new graduate students arrive, and faculty are on campus to teach), we see increases in SMS usage (see Figure 20). Although we do not measure research activity quantitatively, our experience with the general research activity within the IRMACS Centre follows such a pattern. The preliminary evidence presented here suggests that this research activity is at least partially responsible for driving SMS usage.

The second is that SMS usage is driven by technology availability. This is partially suggested by our analysis of the steady increase in SMS usage since the IRMACS SMS rooms became available in January of 2005. This can be seen in Figure 20 by considering each month individually. With few exceptions there is a year over year increase in SMS usage. Given that the SMS rooms are not yet at capacity for either collocated or distributed meetings, it is possible that this growth may continue. We also see some indications that technology availability may have a more direct impact. From mid October to the beginning of December of 2009, one of the SMS rooms was closed for a major renovation and hardware upgrade. This six week room closure essentially removed 17% of the SMS room capability that IRMACS provides (one of the six SMS rooms) and one of the two "workhorse" SMS rooms for supporting research meetings. During this same time period, the year to year increase in SMS usage that one would expect (and as we see in most other months) does not occur (see Figure 20, the rightmost column for October, November, and December 2009).

Note that there could be a range of other contributing factors to these declines. For example, we believe that a reduction in research activity may have also contributed to this reduction in SMS meeting frequency, as a number of research projects that are heavy

users of the IRMACS collaboration infrastructure scaled back their research meeting activity during this time period. In addition, the IRMACS Centre lost one of its key technical support people, so technical support for the SMS infrastructure was also reduced.

It is not yet possible to definitively state that research activity and technology availability are drivers of SMS usage. At the same time, our analysis of IRMACS SMS usage suggests that these may indeed be important factors in the creation and operation of a successful SMS infrastructure. In order to make stronger inferences about SMS usage in the scientific community, more research is necessary.

6.4 What Works and What Doesn't Work?

We often think of IRMACS and its SMS infrastructure as a social experiment. We designed the IRMACS Centre to bring people together both socially and intellectually, and the IRMACS SMS infrastructure has played an important role in accomplishing this. The statistics on SMS usage indicate that distance collaboration is an important part of our computational science research communities work practice. Today, many of our users use the SMS technologies seamlessly. There appears to be an ongoing cultural change in terms of our user's ability and desire to use the SMS infrastructure for distant collaboration. Today, this use often occurs with no need for technical support. The users of the IRMACS Centre expect to be able to collaborate with remote colleagues wherever and whenever they have such a need. In Section 6.3, we performed a quantitative analysis of the usage patterns of the IRMACS SMS infrastructure. In this section, we discuss our qualitative, and often anecdotal experiences in designing, deploying, and operating the IRMACS and WestGrid infrastructure from 2005 to 2009.

6.4.1 What Works

In this section, we discuss what we believe to be some of the successes of the IRMACS SMS infrastructure. We believe that the addition of the IRMACS SMS infrastructure has literally changed the way researchers work at SFU. In our experience, researchers do not use technologies unless there is a clear benefit to their research. Although it is difficult to document quantitatively the value of SMS technologies to the researchers and to the

success of their projects, the level of demand that we see from our user community indicates that IRMACS researchers see a significant value in these technologies.

We believe, as do many researchers [OO00], that “Distance Matters” and that it is important to have our SMS facilities close to the user community. Having an SMS infrastructure as a core technology in a research centre that houses over 250 researchers is an effective mechanism for bringing this technology to the user community. Making use of the IRMACS SMS technologies is a simple and natural extension to a researcher’s typical day at IRMACS. Booking a room for a remote collaboration with colleagues across the country is no different than booking a room for a collocated research meeting at the Centre. We believe that this proximity to the user community is fundamentally important to achieving the levels of use that we see in our SMS facilities.

One of the key successes of the IRMACS SMS infrastructure stems from its support of a comprehensive range of technologies within the physical SMS rooms. Originally, we had planned on using AccessGrid as our primary collaboration tool. We rapidly realized that although AccessGrid provided the most capability in terms of creating an advanced SMS environment, the IRMACS research community was going to ultimately determine the set of software and hardware tools that met their needs the most effectively. Rather than dictate the technology that one uses for collaboration, the IRMACS SMS rooms provide the ability for researchers to create new and dynamic collaboration spaces as required. It is our belief that had we dictated the technology available to our researchers, the SMS usage in the IRMACS Centre would be significantly less than current levels. It is important to note that this does not mean that we allow the collaboration technologies used to be dictated by the remote site. Instead, we consult with our researchers, try to determine what their collaboration needs are, and then try to map that onto a set of technologies that will meet those collaboration needs. Fortunately, our SMS environments have been designed such that most collaboration tools can be used seamlessly in our SMS rooms.

One of the key changes that we have noticed in the IRMACS research community is the ability of the more frequent SMS users to function with a high level of expertise in SMS rooms. These rooms, although designed to be as seamless as possible, are complex technical environments and often require a learning period. Note that this period involves

technological as well as social learning and adaptation. Researchers need to adapt to the technology, but perhaps more importantly they also need to adapt to different social processes. It is possible to mitigate the technological learning process through careful technical design, but the social process is malleable and can only be learned over time. It is our experience that the pervasive SMS infrastructure at IRMACS has accelerated this adaptation. IRMACS researchers are exposed to distance collaboration technologies on a regular basis (through regular remote seminars, attending remote project meetings of other groups, and talking to other researchers who utilize the technology), leading to an understanding and even an expectation that remote collaboration is a standard tool that they can use in their research.

6.4.2 What Didn't Work?

In Section 6.4.1, we portray the way the research community uses the IRMACS SMS infrastructure in a positive light. Indeed, over the last five years, we have seen a dramatic growth in usage. We also have a research community that is rapidly becoming familiar with the capabilities of SMS technologies. Of course, getting to this state has not been without its problems and issues and we would be remiss if we did not discuss these in as much detail.

Building an easy-to-use SMS environment is an extremely difficult task, especially when the SMS environment needs to support a wide range of collaboration tasks. While the IRMACS SMS environments were designed to be flexible, we could not have guessed all of the current uses of these systems, nor understood the limitations of some of the originally selected equipment. In order to adapt to both the ever-changing ways the IRMACS SMS rooms are utilized and the constantly changing software tools and collaboration protocols required by those uses, it has been necessary to constantly update and adapt the SMS systems. Given the constantly shifting requirements of the IRMACS research community, this need is not likely to decrease in the future.

Even if we were able to create an SMS environment that was flexible, powerful, and easy to use, a remote collaboration can break down in many ways. In many cases, much of the technology on which an SMS session depends is outside of local control. This can include the quality of the technology at the remote site (acoustic quality, video quality, etc.), the networking infrastructure that joins the sites, security infrastructure at the

remote sites (e.g. firewalls), and even the familiarity of the remote participants with collaboration technologies. Further, even the definition and terms of success are dependant on the expectations of the researchers and can vary widely based on past experience and the researchers understanding about technical capabilities.

The IRMACS approach to mitigating these problems is to be as proactive as possible in establishing a quality initial collaboration for a research group. By investing time and effort in determining appropriate technologies to use, the quality of network between the collaborating sites, and the familiarity of the researchers with collaboration technologies, we attempt to avoid problems during the SMS session itself. Further, by defining the needs of the collaboration up front, many times the researcher themselves will have a better understanding of the role of technology within the overall goals of the collaboration. While it is not possible to remove these problems completely, our experience indicates that understanding the context of the collaboration is the most effective way of building a successful, ongoing collaboration. Conversely, having a strong negative experience in an initial SMS collaboration can stop an emerging collaboration as quickly as any other problem that might arise.

One of the key obstacles to having a successful collaboration is the change to social interaction that is required in these spaces. Although our SMS environments are technologically sophisticated, they do not reproduce a face-to-face environment. An SMS both presents barriers to the collaboration and at the same time provides new opportunities. It is clear that it is necessary for our users to adapt socially to the environment in which they are working. We have found that the level of adaptation is something that is naturally learned, but is learned differently across different users and for different tasks. Some users become adept at using advanced SMS technologies quickly, while others adapt slowly (if at all). The level to which users adapt to these environments can be quite striking, to the point where we have seen collaborating research groups use components of our SMS environments in ways that they were never intended (see Section 7.3.5 for a detailed example) .

6.5 Discussion

The IRMACS SMS infrastructure has been in operation since early 2005. We have learned an enormous amount in the six years that we have planned, deployed, and

operated this infrastructure. The drastic increase in the use of our facilities indicates that SMS is an important tool to the computational science community. It has also been an excellent opportunity to learn and understand how to support the collaboration needs of this community.

The goal of this chapter was to contribute new knowledge about how researchers use advanced collaboration technologies. To our knowledge, our analysis of the IRMACS SMS infrastructure is unique on two dimensions. First, the study of a purposely designed advanced collaboration infrastructure by a large research community (81 projects and 570 researchers) is unique to this study. Second, the duration over which this analysis is carried out is also unique. Although studies have been carried out that analyze a single project of collaboratory over a long period of time [OEJ+08], none have studied the longitudinal use of collaboration infrastructure by a broad research population similar. Our ability to study how the broad research use has changed over a five year period allows us to draw insights that are difficult or impossible to make without a longitudinal study such as this one.

Our quantitative analysis and anecdotal experiences provide us with several important observations that help to reach our objectives and answer our research questions.

- *Objective 1:* Develop a broad understanding of how scientific researchers collaborate.
  - *Research Question 1: How do collaboration patterns change in the presence of technology?* Our analysis indicates that the size of the research community, the level of research activity, and the availability of technology all have an impact on SMS frequency of use. The increase in SMS usage in IRMACS appears to be driven by the combination of having a critical mass of researchers who could benefit from SMS technologies, the availability of the SMS technologies that support their needs, and the availability of technical support to make the collaborations successful.
- *Objective 3:* Evaluate advanced collaboration modalities and technologies for scientific collaboration.
  - *Research Question 1: How do researchers use advanced collaboration technologies?* Our analysis of the use of the IRMACS SMS infrastructure

suggests that researchers make use of the infrastructure for a broad range of purposes, ranging from large scale, distributed presentations that span more than twenty universities to small research meetings between two people at two sites. We explore two of these scenarios in more detail in Chapter 7 (research meetings) and in Chapter 8 through Chapter 11 (distributed presentations).

- *Research Question 2: How well do those technologies work*? The frequency of use of the IRMACS SMS infrastructure alone (486 distributed meetings in 2009) indicates that the technology works well enough to be useful to a broad range of researchers. At the same time, there is clearly a long way to go before we can claim to support distributed computational science effectively. Users are faced with a steep learning curve, both from a technological and a social perspective, and need to adapt to the technologies. Our experience suggests that making an SMS meeting work right the first time is critically important to a successful, ongoing collaboration. Once a research group becomes familiar with the technology, our analysis also suggests that distance collaboration can rapidly become a natural part of a research groups work process. In order to facilitate this, it is critical to have an SMS infrastructure that is easy to use and well supported. The wide range of use of the IRMACS SMS infrastructure suggests that no single technology will meet all researcher needs, and therefore an SMS environment should be both flexible and extensible.

We believe that it is the combination of the critical mass of research activity at IRMACS, the availability of the technology, and the high level of technical support that that have resulted in the dramatic increase in SMS usage at IRMACS and SFU. Through the continuation of the IRMACS efforts, we believe that the computational science community's use of SMS technologies is only beginning to evolve. Although more research needs to be performed to determine the specific impacts of each of these factors, our experiences with the operation of the IRMACS SMS infrastructure from 2005 – 2009 suggests that all of these factors need to be taken into account when building, deploying, and operating an SMS infrastructure for the computational sciences.

7 Artifact-Centric Collaboration – An Ethnography

In Chapter 6, we explore how the collaboration usage patterns of a broad scientific research community are impacted by a carefully designed and well supported collaboration infrastructure. Chapter 6 provides us with a high-level view of the frequency of use of advanced collaboration technologies within a large research centre and suggests that technology can have a dramatic impact on the usage patterns of such a technology. It also provides us with a basic understanding of the types of collaborative meetings that take place (research meetings, seminars, etc.). Unfortunately, it does not inform us on how the technology is used. What do the researchers really do in those meetings rooms? What technologies do they use, and how? How well do the technologies work? Is collaboration around digital artifacts really important?

In this chapter, we provide evidence that begins to answer such questions. Rather than provide a broad, high-level view of collaboration use across many projects (like the one provided in Chapter 6), this chapter carries out a detailed analysis of how technology is used in specific research scenarios. We are interested in the impact that distance has when remote collaborators are working together with digital artifacts that are complex in form, such as data that results from complex scientific simulations. In particular, we explore the importance of gesture in collocated and distributed scientific collaboration. This chapter helps to reach our research objectives and answer the following research questions:

- *Objective 1: Develop a broad understanding of how scientific researchers collaborate.*
  - *How do collaboration patterns change in the presence of technology?*
- *Objective 2: Develop a deep understanding of how scientific researchers interact with digital artifacts when they* collaborate.
  - *What role do digital artifacts play in scientific collaboration?*
  - *What information is lost when such collaboration takes place at a distance?*
  - *What communication channels are used to encode information during artifact-centric collaboration?*

- *Objective 3: Evaluate advanced collaboration modalities and technologies for scientific collaboration.*
  - *How do researchers use advanced collaboration technologies?*
  - *How well do those technologies work?*

This chapter presents the results from a longitudinal ethnographic study of a group of collaborating scientific researchers. We observed a single research group during its regular research meetings, performing over 18 hours of observations spanning a five-month period. A coding scheme for artifact-centric interactions was developed and applied to those meetings that involved significant artifact-interaction. We present a high-level analysis of the type and structure of the meetings held by the research group, as well as a detailed analysis of a number of meetings where artifact interaction is prominent.

7.1 Studying artifact-centric collaboration

7.1.1 Observational study

The goal of this research is to gain a better understanding of how scientific collaborators interact with digital artifacts during meetings. We are interested in observing both collocated and distributed research groups. This is important, as we want to understand how collocated researchers work with digital artifacts and what information is lost when these collaborations occur at a distance. We attempt to gain this understanding using naturalistic, longitudinal observational studies in the field. We observe a single research group during normal work meetings in both collocated and distributed settings. Ethics approval for this study was obtained from the University of Victoria Human Research Ethics Board. Meetings are recorded on video tape for later analysis. A single observer records the meeting, manipulating the camera to focus on specific activities as appropriate. The observer also takes notes on any interesting events that occur during the meeting. The observer does not participate in the meeting.

7.1.2 Coding

Analysis of the data gathered from these studies used an open and emergent coding scheme. Since the study is exploratory in nature, it was decided that the types of events in which we were interested should emerge from the actions of the researchers. This approach was taken for two reasons. First, few, if any, naturalistic, longitudinal studies of

a single scientific research group have been reported in the literature. Thus, we have no direct, domain specific literature on which to base our coding schemes. Second, the naturalistic observation of users in a technologically sophisticated (e.g. like the SMS environments described in Chapter 6) production meeting room (i.e. not a CSCW research environment) are also rare. Although the specific codes used in this study emerged as the study proceeded, the high-level structure of the coding scheme was based on a gesture coding schema from the Department of Linguistics at Goteborg University in Sweden [Cer02] with influence from the work of Tang and Leifer [TL88], Bekker *et al.* [BOO95], and McNeill [McN92, p. 377].

We used two main coding categories, a *structural* code and an *action* code. Structural codes marked moments in the meeting where the structure or phase of the meeting changed. These phases can be mapped to task level categorizations as presented in the CoGScience Framework (see Section 4.4.1). Two common meeting phases are **description** phases and **discussion** phases. **Description** phases of a meeting occur when one person is describing something to the group (some data, a mathematical model, or a paper he/she had read) or giving a presentation on his/her research. These are **performance** tasks from the CoGScience Framework. **Discussion** phases are interactive phases between two or more individuals and would commonly consist of either **planning** (research project planning), **creative** (coming up with a new mathematical model), **intellective** (solving a problem), or **decision** (agreeing on a research approach) tasks from the CoGScience Framework.

Action codes annotate actions made by participants in the meeting. We use a wide range of action codes, including **verbal utterances**, **gestures**, **body language**, **facial expressions**, and physical actions (such as writing or typing). These actions can be mapped to the **channel** level in the CoGScience Framework. Each action code has several subcodes, providing a mechanism for refining the analysis. For example, a verbal utterance might be classified as a statement, a question, a response to a question, verbal feedback, or referring to an artifact (similar to that defined by [Cer02]). Gestures are coded in a similar way, with coding differentiating between gestures that point at physical objects, gestures that point at digital artifacts on the screen, gestures that refer to a person, and gestures that are used for emphasis ("it was this big" and indicating size with your

hands or general "hand waving"). This differentiation is similar to those provided by McNeill [McN92], Tang and Leifer [TL88], and Bekker *et al.* [BOO95].

All coding of events were performed through post-meeting analysis of the video taped recordings of the meetings. The author analyzed and coded all meetings. The emergence of event codes was a subjective process and was based on the emergence of themes relevant to artifact interaction that were witnessed in early meetings. After several meetings, the coding scheme reached a steady state and refinement no longer occurred. It is this coding scheme that is used in the analysis presented here. Each code includes an event identifier, a primary event code, a secondary event code, the time the event took place, the visual stream used to communicate the event, whether there was a problem with the communication of the event, and any additional comments about the event.

Although the process of creating the coding scheme was subjective, the application of the final coding scheme to a specific meeting is relatively mechanical. If an utterance is made, the type of utterance is noted and coded as described above. The same is true when a gesture is made. This straightforward application implies that both the coding process and the coding scheme can be used as a tool to analyze a variety of artifact-centric collaborations. A subset of the codes used in this study (gesture and utterance codes), as well as an example coded segment from one of our observed meetings, is given in Appendix 15.4.

7.1.3 Emergent high-level gestural interactions

After performing a detailed analysis of the coding for a number of meetings (including M3 and M4), we discovered that the low-level events that were coded could be grouped into composite events that had meaning above the individual events. Question and answer pairs, gesture and utterance pairs, and gesture, utterance, and action triads are all potentially interesting composite interaction events. In addition, we differentiated between gestural events when they occurred physically (someone physically points), on the computer (someone points with the mouse), or with the Smartboard. We then analyzed this information over time to determine structure and flow of the meeting and to expose themes and patterns in how users interact with digital artifacts.

We have created criteria for composing events into high-level composite events. In particular, we have defined two types of important high-level artifact interaction

communication events. We define these as **explicit** and **implicit** artifact communication events. An explicit artifact communication event occurs when the following criteria are met:

- An utterance event occurs;
- A gesture event occurs;
- The utterance and the gesture events are generated by the same individual;
- The utterance and gesture events occur at approximately the same time;
- The utterance refers to an artifact;
- The gesture refers to an artifact; and
- The utterance is deictic in nature (that is, the utterance makes an explicit reference to an artifact, such as *"this is the answer"* while pointing to a number in a cell in a spreadsheet, for example the number 42).

We call such a communication event an explicit artifact communication event because the artifact that is the referent of the communication cannot be implied from the deictic utterance and requires the ***explicit*** gesture for the communication event to be interpreted correctly. We often refer to such a pairing of a pointing gesture and a deictic utterance as an explicit artifact gesture.

Implicit artifact communication events are similar to explicit communication events except that the utterance is not deictic in nature. That is, only the last criterion is different in the above list. For example, an utterance of *"the answer is 42"* while pointing to the number 42 in a spreadsheet would be an implicit artifact communication event. We call this an ***implicit*** event because the referent artifact can be ***implied*** from the utterance. Implicit artifact events have enough meaning implicit in the utterance to identify and locate the referent artifact without the gestural component. That is, the artifact and its location can be understood from the utterance without the location being directly expressed. This makes the gesture at least partially redundant.

Our exploration of distributed, artifact-centric, scientific collaboration suggests that differentiating between these two types of artifact interaction events is important. In order for an individual to understand an explicit artifact event, both the utterance and the gesture must be communicated to other individuals (in particular, to remote participants in a distributed collaboration). Without the gestural component of the event, an explicit

artifact event has little or no meaning. Implicit artifact events have enough meaning implicit in the utterance to identify and locate the referent artifact without the gestural component. Given that we are interested in the impact of distance on artifact-centric collaboration, it was deemed that this distinction was of critical importance. We need to understand the frequency and usage patterns of these types of high-level communication events.

It is important to note although it may be possible to identify the referent artifact without gesture in an implicit artifact event, the gestural component may still be critical to the communication. For example, in a spreadsheet that has more than one 42, the utterance *"the answer is 42"* does not have enough implicit meaning to disambiguate which of the 42s is the referent artifact. It is clear that we need to consider both implicit and explicit artifact events in our exploration of distributed artifact-centric, scientific collaboration.

Three other types of composite gesture interaction events were also identified. We call these artifact manipulation events, atomic artifact gestures, emphasis gestures, and spatial gestures. Artifact manipulation gestures occur when an individual manipulates an artifact (a spreadsheet). Atomic artifact gestures (these are in fact not composite artifact events) are gestures that refer to an artifact but are not accompanied by an utterance. Emphasis gestures are general kinetic gestures that do not refer to a specific artifact (hand-waving). Spatial gestures contain information of a spatial nature but do not directly point at an artifact or object ("it was this big").

Note that there is a tight coupling between the artifact and gesture events described above (implicit/explicit, manipulation, emphasis, and spatial events) and the human gestural communication channels described in the CoGScience Framework (**pointing**, **manipulation**, **kinetic**, and **spatial** gestures). The focus of our analysis is on explicit and implicit artifact events, primarily because the majority of gestural interactions fall into these two categories (see our analysis below). The artifact gesture events that do not fall into these two categories are typically artifact manipulation gestures, atomic gestures, or spatial artifact gestures.

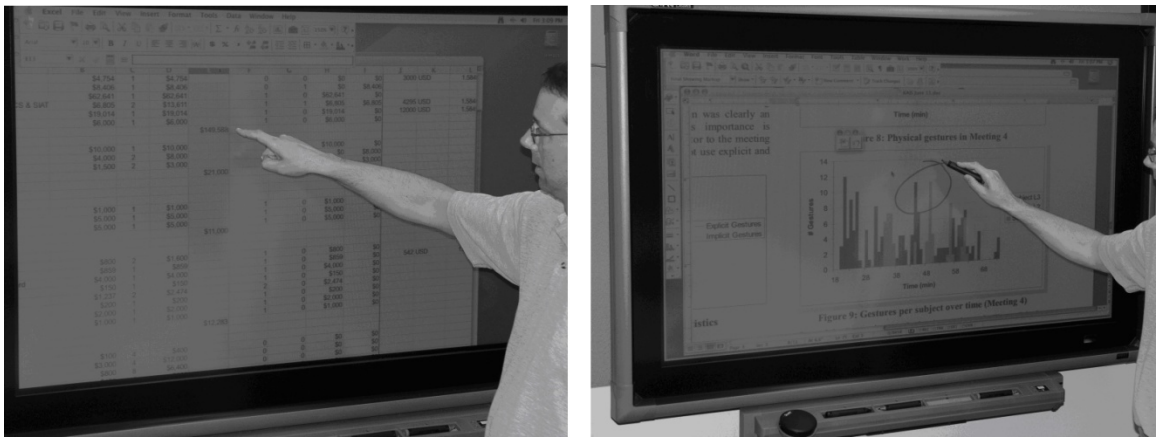


Figure 22: Physical pointing (left) and Smartboard (right) gestures.

Example explicit artifact interactions that might occur are a participant highlighting a single cell in a spreadsheet with the mouse combined with the utterance "…this is the number entering the system" or a participant circling a feature in a graph using one of the Smartboard pens combined with the utterance "…this is the time where this parameter peaks…" (the right image in Figure 22). An example implicit artifact event would be a participant physically pointing at a number from the output of a computational simulation combined with the utterance "…42 is the output for the first phase…" (the left image in Figure 22).

7.2 Ethnography Study Description

The ethnographic study presented here focuses on the naturalistic study of an active research group in its traditional work environment. Our goal was to study both collocated and distributed collaboration, both of which the research group performs on a regular basis.

7.2.1 Subjects

Our ethnography was a longitudinal study of a small research group of computational scientists. The group consisted of 14 members (six females and eight males) and met once or twice a week to work on a variety of projects. Projects revolved around the modeling of complex processes and phenomena. Meetings varied in their purpose and content, including planning sessions for new projects, open discussion sessions, presentations to the group, and focused research meetings. Several of the research group members had worked on research projects before and were therefore familiar with each

other. Most of the researchers involved in the project were acquainted with each other before the research project started, although there were some new members in the group. The group members had been working on the project for several months before our observations began.

The research group uses modeling and computational simulation as tools for understanding and optimizing complex systems. During our observations, the research group was attempting to model and understand the functions of a complex, real world system. The goal of the group's research was to gain a better understanding of how the system functioned, identify bottlenecks in the system, explore what-if scenarios by changing the operation of the system, and ultimately optimize the system to meet target goals. The research group, starting with data that had been gathered about the system over time, created mathematical models of the system that could then be used to understand, simulate, and optimize the system. The models were validated by running the models and comparing the results to the observed data. In addition, multiple models of the system (using different modeling approaches) were created and validated against each other.

Much of the group's work revolved around digital artifacts in various forms. This included papers that were being presented or discussed, spreadsheets that contained the results of computational simulations, graphs or other visualizations of the simulation and modeling results, the source code of the computational simulation itself (for development or explanatory purposes), and digital sketches of brainstorming concepts for project planning and development.

7.2.2 Technology environment

Group members traveled often during the observational period, and therefore attendance at group meetings from remote locations was often necessary. The group used a variety of collaboration tools. An audio communication channel was always used, and was provided using either an analog phone or IP (network) based audio collaboration tools (Skype [MR06], iChat [Car03], or AccessGrid [COP+00]). Video of remote group members was sometimes used, using either iChat or AccessGrid.

Shared documents between remote participants were provided using Virtual Network Computing (VNC), a desktop sharing tool [RSW+98]. Using VNC allowed a remote

participant to see another participant's desktop, including mouse motion and digital mark-up using the Smartboard. It was also possible for remote collaborators to take control of the remote desktop and manipulate artifacts as if they were on their local desktops. Artifacts, and how they were manipulated, are discussed in more detail in Section 7.1.3.

Figure 23: A typical advanced meeting room used during the study

The group used sophisticated rooms like those described in Section 3.4 (pictured in Figure 23) for the collocated members. Remote users typically used desktop or laptop computers. Rooms typically had one or more plasma displays mounted on the wall, with Smart Technologies touch screen overlays (Smartboards) that allowed users to draw directly on the screen. The rooms had permanent computers that drove the two displays, as well as the ability to plug a laptop into a display. The permanent computer was connected to, and controlled, the Smartboards. The rooms also contained sophisticated AV components, including an acoustic echo canceller (providing good quality full duplex audio) and video cameras for sending video streams to remote collaborators.

It is important to note that the meeting rooms the research group used for meetings were the same meeting rooms that they used before our observations began. That is, the work environment did not change to facilitate our study. The use of sharing documents through desktop sharing (VNC) and the use of Smartboards were new to the group as a whole and were also new to many of the individual researchers. One of the researchers was relatively familiar with the technology, including desktop sharing and the Smartboards, and was typically responsible for coordinating the setup of the meetings.

7.2.3 Observed meetings

Our study included the observation of eleven meetings of approximately one and a half hours each over a five month period, for a total of approximately fifteen hours of raw data. The meetings ranged in topic from casual discussions through to intense analysis of computational models and the modeling results. Six of the eleven meetings we observed involved significant artifact interaction. All of these meetings were coded using the coding scheme described in Section 7.1.2. The focus of the coding was on events that involved interaction with artifacts. For example, in meeting phases where there was no artifact interaction, utterances were not coded. All gestures and utterances that referred to artifacts or objects were coded, including gestures that referred to objects or artifacts indirectly (such as statements "it was this big").

In the analysis below, we explore three meetings (out of the eleven meetings that we observed) that were of particular interest. Meeting three (M3) and meeting four (M4) were data and model analysis meetings and were therefore highly artifact-centric. In addition, M3 was a distributed meeting while M4 was collocated, providing an interesting contrast to how the users interacted with each other. Meeting eleven (M11) involved a presentation by one of the researchers, followed by a discussion about the topic of the presentation. M11 had two remote participants. We chose M3, M4, and M11 because of the particular features of the collaboration that they portray. M3 and M4 are very similar meetings but one is distributed and one is collocated. M11 is a distributed meeting with extensive use of gesture, allowing us to analyze the effectiveness of how that gesture is communicated to the remote participant. Although we don't analyse all meetings in depth, we occasionally discuss specific observations from other meetings. A more detailed description of the meetings from our ethnography is given in Appendix 15.5.

7.2.3.1 M3 Description

M3 was a distributed meeting with one of the group members joining the meeting from overseas (from a hotel room) with the remainder of the group (six of them) participating from the room shown in Figure 23. The meeting lasted one hour and fifteen minutes. The main topic of the meeting was the discussion of the data set and a mathematical model of the system. The model was instantiated as a computer simulation and produced numerical results. Several documents were used during in the meeting including a spreadsheet and

the code for the simulation itself. The spreadsheet contained the raw data gathered from the real system, data output from the computational simulation, and several visualizations of the data being discussed (in the form of graphs). The computer mouse, the Smartboard, and physical pointing gestures were used to interact with artifact interactions during this meeting.

From the perspective of the CoGScience Framework, there were two **sensory streams** used in this meeting. There was a moderate **fidelity aural stream**, utilizing an overseas phone connection to a hotel in Europe. There was a **high fidelity** (1024 x 768 pixels) application **visual stream** of the computer desktop (using VNC) sent to the remote collaborator. This allowed the collaborator to see any application running on the computer as well as any interactions that were performed using the mouse or the Smartboard. There was no visual stream that allowed the remote participant to see the other participants in the room. Nor was there a visual stream that allowed remote participants to see physical gestures (made with the hand) in the context of the task space (Buxton's reference space [Bux09]).

7.2.3.2 M4 Description

M4 took place five days later and was a similar meeting in basic structure to M3. The goal of the meeting was to explore further the system being modeled and to validate the model that was being developed. One of the other participants had developed a second, independent mathematical model for the system, and this model was also explored in the meeting. The main difference between M4 and M3 in terms of meeting composition was that all participants in M4 were collocated. One additional member joined the group and the remote user from M3 was now on site. Both a laptop and the built-in room computer were used during this meeting, with the mouse, Smartboard, and physical gestures used to interact with artifacts on the screen.

M4 was a particularly interesting meeting, as it started with a relatively sedate presentation about the data and simulation being considered (involving one person), went through a discussion phase where a potential problem in the model was identified (with three people actively involved), and then transformed into an intense and interactive problem solving phase involving most of the members of the group. Eventually, the problem in the computer model was found and the problem was solved. As these phases

progressed, interest, engagement, and excitement gradually increased. We analyze this progression in some detail in Section 7.3.6.2.

7.2.3.3 M11 Description

M11 was a very different meeting from M3 and M4. M11 focused on the discussion of two papers that were relevant to the group's research. The papers presented models that the group was considering integrating into their research. The papers were mathematical in nature, and much of the discussion revolved around the formulas and figures that were contained in the papers. The two papers were presented by two different participants, with both presenters at the local site. There were four collocated and two remote participants. The paper was viewable on one of the displays in the room, and participants interacted with artifacts on the screen using the mouse, the Smartboard, and physical pointing gestures.

From the perspective of the CoGScience Framework, this meeting was very similar to that provided in M3. There were two **sensory streams** used in the meeting. There was a moderate **fidelity aural stream**, utilizing Skype between the local site and the two remote sites. There was a **high fidelity** (1024 x 768 pixels) application **visual stream** of the computer desktop (using VNC) sent to the remote collaborator. This allowed the collaborator to see any application running on the computer as well as any interactions that were performed using the mouse or the Smartboard. There was no visual stream that allowed the remote participants to see the other participants in the room. Again, there was no visual stream that allowed remote participants to see physical gestures (made with the hand) in the context of the task space.

7.2.3.4 Other Meetings

There were eight other meetings recorded as part of our ethnographic study. Six of the meetings were collocated and two of them were distributed. Of the distributed meetings, one of them had significant artifact interaction while the other had almost no artifact interaction (primarily a discussion with no use of digital artifacts). Of the six collocated meetings, all meetings contained artifact interaction of some type with two other meetings having extensive artifact interaction similar to M3, M4, and M11. Our analyses

of the other meetings provide us with similar results to those presented above but we do not provide a detailed analysis of these meetings here.

7.2.4 Focus Group

At the end of the observational period, a focus group was held with the study participants. The goal of the focus group was to explore the researcher's experiences with scientific collaboration, with a focus around their use of data during their collaborations as well as their use of collaboration technologies. The focus group session took 90 minutes and was facilitated by the author. The focus group involved a pre-planned set of questions that requested the participant's feedback about the importance of collaboration (in general) to their research group as well as on the importance of collaboration around data or documents (see Appendix 15.3 for the focus group script). In order to spark discussion about collaboration scenarios, participants were asked to give examples of how they shared data/documents when collocated and when distributed. Participants were also asked to discuss how they thought their collaboration patterns had changed during the study period. We use the discussion generated in the focus group in our analysis below.

7.3 Analysis and Results

Our analysis proceeds as follows. We first analyze the structure of the various meetings we analyzed as part of this study (Section 7.3.1). We then analyze the amount of artifact interaction and gesture use throughout the meetings (Section 7.3.2) and consider the impact that distance has on those gestures (Section 7.3.3). We then analyze the impacts of a number of other factors on the meetings we analyzed, including individual differences (Section 7.3.4), how participants learned and adapted over time (Section 7.3.5), and the interactions between physicality, engagement, and gesture (Section 7.3.6). Lastly, we distil these analyses into a set of coherent research results that address the relevant research questions that this study is designed to answer (Section 7.4).

We encapsulate our analysis in the context of the CoGScience Framework. In particular, we consider the type of **task** being carried out and the **task characteristics** as variables that have a fundamental impact on how researchers interact with digital artifacts. From a CoGScience perspective, we consider a number of task **characteristics**,

including the **nature of the material** (artifact or non-artifact related) being dealt with during a meeting is fundamental to our analysis. In addition, the level of **coupling**, **exploration**, **creativity**, **difficulty**, and **complexity** are all relevant task characteristics that we consider in the context of our analysis.

7.3.1 Meeting structure

Our analysis of meeting structure primarily contributes to our research objective of trying to "Develop a broad understanding of how scientific researchers collaborate." That is, we explore meeting structure to understand the process of scientific collaboration. Our findings from this analysis are as follows:

Finding 1 Research meetings proceed in phases and each phase often incorporates a different task function (e.g. describe versus discuss). Researchers switch between phases rapidly.

Meeting structure across the meetings analyzed varies dramatically, depending on the task being carried out and the characteristics of the group. Using the structural codes defined in our coding scheme (see Section 7.1.2), we analyse each of the meetings. From a CoGScience perspective, research meetings tend to consist of either **planning tasks** or **intellective tasks**. **Planning tasks** consist of either description phases where a researcher is performing a **reporting function** on the status of the research project or discussion phases where the group is performing **coordination**, **mediation**, or **execution functions** relevant to advancing the project. **Intellective tasks** also tend to fall into description and discussion phases, where the description phases consist of researchers performing **explanatory** or **reporting** functions about research findings and the discussion phases where researchers are performing **exploration** and **idea generation functions**. Statements from the focus group indicate the researchers perceive a similar process:

> *"…the vast majority focuses on one person preparing a topic or presentation, and then telling the story to other people through a PowerPoint presentation, using some other tools like papers. Then, a smaller proportion of the interaction is where we actually work on something together, we are creating something from scratch together or working on a previously prepared document, but where all of us are inputting information into that."*

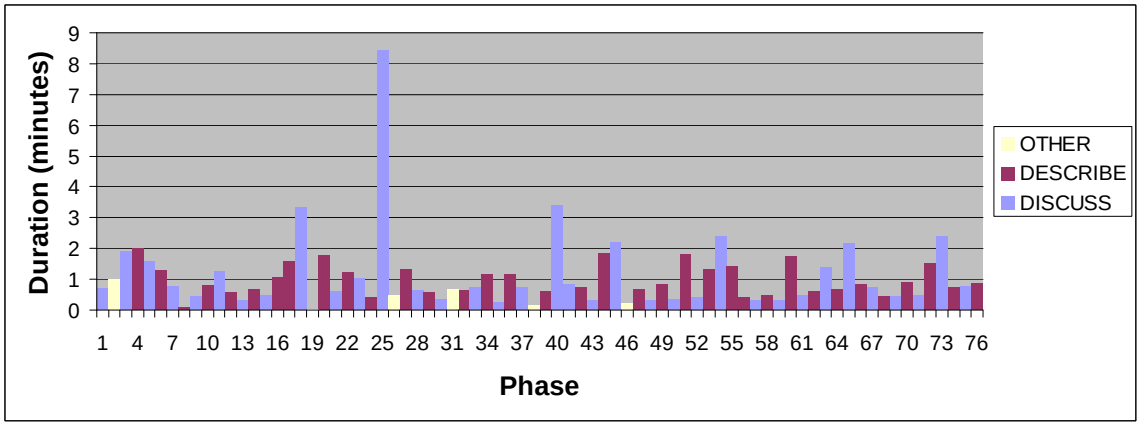


Figure 24: Phase durations for Meeting 4

Like many things, the phases of a meeting appear to span a continuum. We analysed only those meetings where the **nature of the material** was artifact centric. M4, the most dynamic and interactive meeting had 77 meeting phase changes in an 80 minute meeting. This meeting was **tightly coupled**, **exploratory**, **creative**, and **difficult**. In particular, it is this **tight coupling** that is reflected in the number of meeting phase changes. The longest phase in this meeting was a discussion phase of 8.42 minutes with the average phase duration of 1.07 minutes (see Figure 24).

In many parts of M4, the presenter would describe a data set or model parameter for 30 seconds and the group would immediately ask questions. Asking the question moves the meeting from a description phase to a discussion phase. The meeting remains in a discussion phase until the presenter starts describing another element in the data or model. Given that there are 77 meeting phases in an 80 minute meeting, the meeting is clearly very dynamic. Some description/discussion phase changes were as brief as one minute, typically consisting of a description and one or more clarifying answers, before the presenter continued the description.

M3, which was a distributed meeting, was slightly less interactive with 48 meeting phase changes over a 94 minute meeting (31.35 minute maximum phase duration, 2 minute average phase duration). Compared to M4, this meeting was also highly **exploratory** in nature but had a lower level of **coupling** (less intense questions and discussion). M11, which was also distributed but also more presentation based (that is, participants were presenting paper summaries), had only 24 meeting phase changes during the 94 minute meeting (16.42 minute maximum, 4.09 minute average). This

meeting was not tightly coupled, with a relatively small number of phase changes and some fairly long discussion phases.

7.3.2 Artifact Interaction and Gestures

Our analysis of artifact interaction and gesture use helps us to answer several of our key research questions:

- *What role do digital artifacts play in scientific collaboration?*
- *What communication channels are used to encode information during artifact-centric collaboration?*
- *What communication channels are used to encode information during artifact-centric collaboration?*

Focus group participants indicated that the sharing of artifacts (spreadsheets, presentations, or papers) was critical to their work process. For example, one participant stated that

> *"…this general idea, there is some kind of document on the screen, it is the focus of discussion, and somebody is leading other people through it, through that document, and VNC is making that accessible to people … is 90% of what we do."*

In particular, the ability to interact with artifacts using gestures remotely was deemed critical.

> *"What we were doing here, with L1, was very much back and forth, I [remote participant] was commenting, L1 was asking questions, L2 was saying what about this part, lets look at this part, it was extremely interactive on all levels. I HAD to have mouse control there, if I did not have mouse control, or at least the ability to point is what I am talking about when I mean control, it wouldn't work…"*

> *"You know what was really valuable about that was not so much the ability to move the screen, that was OK, what was really essential was for me to be able to highlight a part of the screen and say this is the cell I am talking about right now. I can do that remotely and we can both see exactly which part I was talking about."*

All three meetings had extensive artifact interaction and gesture use. The frequency for explicit and implicit artifact gesture events for M3, M4, and M11 are given in Figure 25 through Figure 27 respectively. These figures show the number of artifact events that

occurred over the span of a single minute in each meeting. For example, Minute 21 in Figure 26 represents the one-minute period of M4 from the end of minute 20 to the end of minute 21. As the figure shows, during this period 5 explicit and 6 implicit artifact gesture events occurred, giving a total of 11 artifact interaction events for that time period. It is worth noting that some graphs do not start at Minute 1. For example, Figure 25 starts at Minute 40, as no gestural interactions occurred before that time in M3.

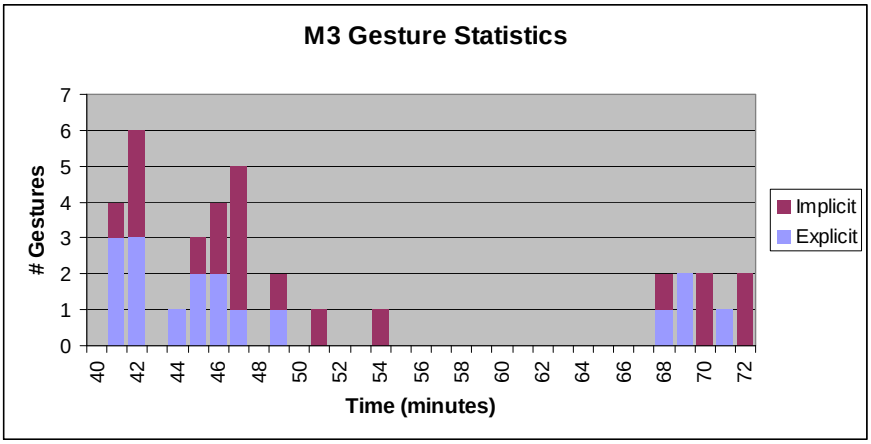


Figure 25: Meeting 3 (M3) explicit and implicit artifact gesture events

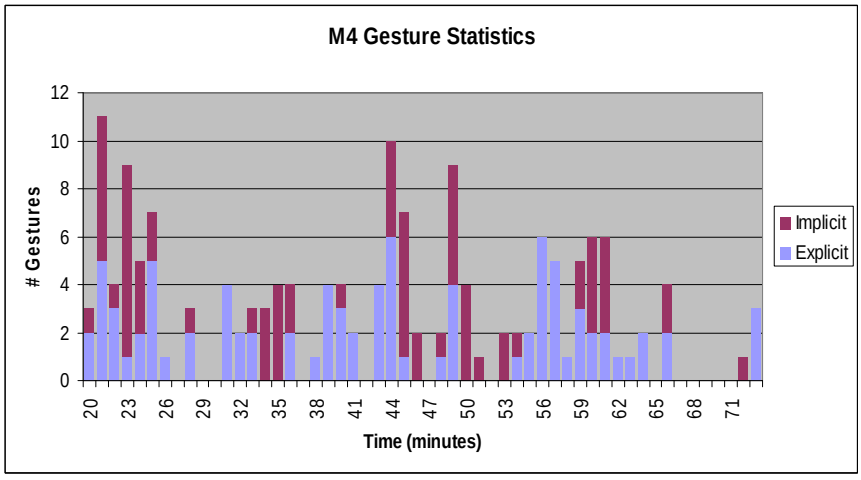


Figure 26: Meeting 4 (M4) implicit and explicit artifact gesture events

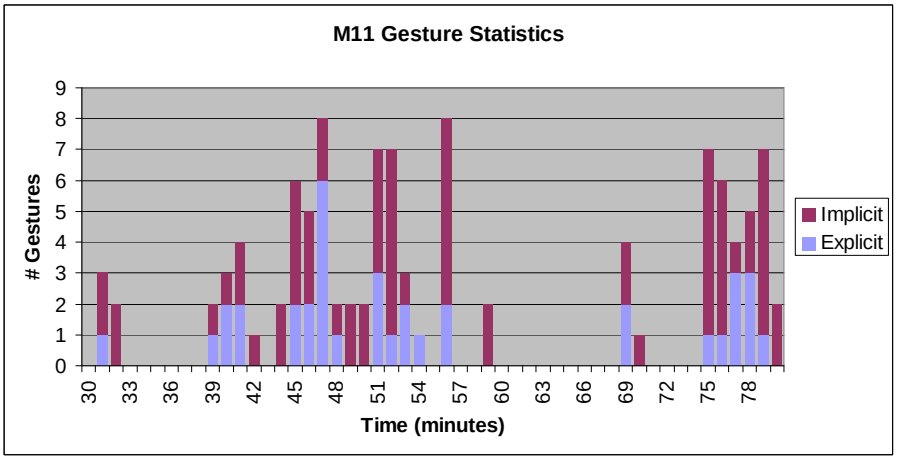


Figure 27: Meeting 11 (M11) implicit and explicit artifact gesture events

Finding 2 Artifact gestures are used frequently in both collocated and distributed scientific collaboration.

Minute 21 in M4 (Figure 26) gives the highest number of gesture events per minute out of all three meetings, with artifact gesture event frequencies above 6 gestures per minute (an artifact gesture every 10 seconds) are common. M3, M4, and M11 have maximum artifact gesture frequencies of six, eleven, and eight artifact events per minute respectively. Other meetings we observed (but are not presented in detail here) result in similar artifact interaction frequencies.

In general, our artifact event frequency analysis shows that artifact gesture is frequently used in scientific collaboration, and in particular in artifact-centric, scientific collaboration. Note that these artifact gestures occur in both collocated (M4) and distributed meetings (M3, M11). Recall that each artifact gesture event noted in the figures above is a gesture that points to an artifact on the screen accompanied by an utterance about that artifact (either explicit or implicit artifact events as defined in Section 7.1.3). These artifact events are very specific to the artifacts involved. Our study indicates that scientific collaboration has similar levels of artifact interaction as other artifact-centric collaboration domains [BOO95, TL99, Tan89]. For example, Bekker observed up to 14 gestures per minute in a collocated design meeting. One important difference worth noting is that our artifact gestures occur at relatively high frequency levels in both collocated and distributed meetings.

Finding 3 Artifact gestures are not used in all phases of a research meeting.

One pattern that is important to note is that although artifact gestures are used frequently in research meetings, there are significant parts of the meeting where there are no artifact gestures (see Figure 25 through Figure 27). This implies that although artifact gestures are used frequently in scientific collaboration, they are not used all the time. This is not surprising, given that in the previous section it was shown that meetings are highly structured and proceed in phases. A project planning phase (scheduling the next research meeting for example) where the **nature of the material** is non-artifact related is unlikely to consist of complex artifact interactions.

Finding 4 Artifact gesture frequency is often high when a single researcher is describing a complex topic that involves artifacts (a loosely coupled, description task).

Two factors that appear to contribute to high artifact gesture event frequencies are the **function** being performed and the **coupling** with which the task is being carried out. Artifact gestures appear to be frequent in **description** phases where a single presenter is describing a complex set of data. This often implies that the presenter is rapidly pointing at a wide range of artifacts on the screen. In addition, phases in which gesture is prominent are often (but not always) those with low **coupling**. That is, phases when a presenter is describing data, a diagram, or a figure to the group without extensive interaction often contain frequent artifact gesture events. Low **coupling, description** phases with high artifact gesture event frequencies are prominent in parts of M11 (Minute 44 to 48, Minute 51 to 56, and Minute 75 to 80 in Figure 27).

Finding 5 Artifact gesture frequency is often high when multiple researchers are involved in the discussion of complex scientific artifacts (tightly coupled, discussion task).

It is important to note that loosely couple description phases are not the only phases in which artifact gesture events occur frequently. M4 is the most dynamic meeting of the three, and this is partially reflected in the number of total artifact gestures used (160 artifact gestures over a 75 minute meeting, or on average 2.1 gestures/minute). Recall that in our description of M4 (see Section 7.2.3.2, Section 7.3.1, and Appendix 15.5.2) the **nature of the material** is artifact-centric, **tightly coupled** (user interaction), **exploratory** (exploring a new data set and new computational model), **creative** (creating new models), and **complex** (the system being modelled is complex). In addition, there is **urgency** (the group has a deadline approaching), **competitiveness** (multiple models from

different group members are being considered), and **conflict** (there is an inference that one of the models is wrong) in the meeting. Although there are times where there are no artifact gestures being used, there are also significant portions of the meeting where gesture use is prominent. We discuss the interaction of these task characteristics in more detail in Section 7.3.6.

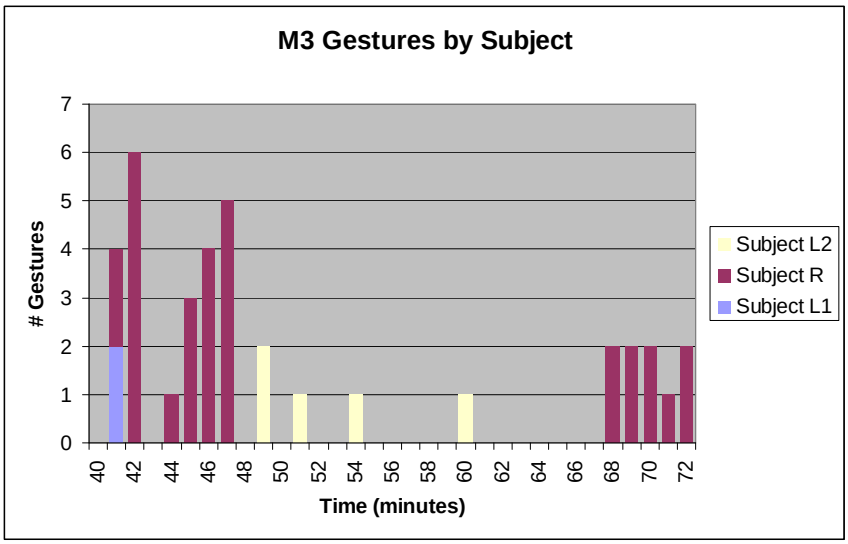


Figure 28: Number of artifact gesture events by participants in M3

Finding 6 During a given phase of a research meeting, artifact gestural interaction is often performed by a single researcher.

Contrast this to M3, a similar meeting in terms of high-level **task**, but with significantly different artifact interaction statistics (36 artifact gestures over a 60 minute meeting, or on average 0.6 gestures/minute) and **task characteristics**. The **nature of the material** is still artifact-centric, but the meeting is much less **tightly coupled** (fewer phase changes). In addition, recall the meeting M3 is a **distributed** meeting, with one participant joining from Europe. In addition to fewer phase changes, the artifact events generated by participants are relatively coherent, with artifact gesture events typically being made by a single participant over an extended period of time. Figure 28 show this pictorially, with the remote subject (R) making the bulk of the artifact gestures (between Minute 41 - 47 and Minute 68 - 72). This lack of temporal artifact interaction between participants (only one person pointing) further indicates that the **coupling** in this meeting is relatively low.

M11 is also less **tightly coupled** than M4, with the fewest phase changes of the three meetings (24 in M11 versus 77 and 48 in M4 and M3 respectively) but a relatively large number of artifact gesture events (106 artifact events in 94 minutes, for an average of 1.13 artifact gestures/minute). The papers being discussed in this meeting present **complex** topics, and most of the artifact gestures generated during this meeting are performed to refer to specific terms in complex mathematical formulas. Thus, the gestures appear to be used to help deal with the **complexity** of the topic being presented and disambiguate the utterance being made. Most of the artifact interaction is performed by a single individual (approximately 95% of the artifact gestures), and as described above, the higher frequencies of artifact gesture events occur in description phases of the meeting (Minute 44 to 48, Minute 51 to 59, and Minute 75 to 80 in Figure 27).

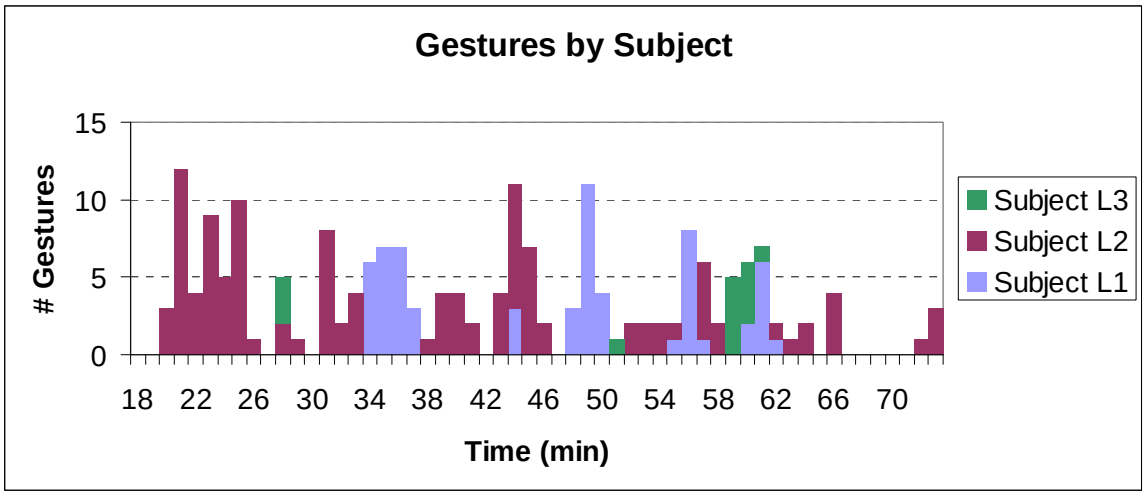


Figure 29: Gesture by subject for Meeting 4 (M4).

Finding 7 During a given phase of a research meeting, artifact gestural interaction is sometimes performed by a multiple researchers in a short period of time.

In contrast, we again look at M4. This meeting is a highly dynamic, tightly coupled meeting, with significant amounts of gestural interaction. In this meeting, a large number of the phases are similar to those in M3 and M11 in that a single person is generating the bulk of the artifact gesture events (e.g. Minute 20 – 25 and Minute 34 - 37 in Figure 29). There is also a phase of the meeting where three of the participants generate a number of artifact gesture events in quick succession (Minute 51 – 64 in Figure 29). This is the intense, problem solving portion of the meeting (described in Section 7.2.3.2), where most of the group members are actively involved in identifying a key problem in the

computer model under consideration. It is clear that in some phases of some meetings, multiple people interact with digital artifacts in a short period of time.

7.3.3 Impacts of distance

Two of our primary research questions focus on determining *What communication channels are used to encode information during artifact-centric collaboration?* and *What information is lost when such collaboration takes place at a distance?* In order to answer these questions, we must answer two other research questions, that is *How do researchers use advanced collaboration technologies?* and *How well do those technologies work?* In this section, we consider these research questions.

In comparing M3 and M4, it is tempting to look at artifact gesture frequency and hypothesize that collocated meetings such as M4 (Figure 26) have more artifact interactions than distributed meetings such as M3 (Figure 25). Given that these two meetings are very similar in nature (exploration of a complex data set and the related computational model used to simulate the system that produced the data), this is a logical comparison to try and make. Through the application of the CoGScience Framework and our analysis presented above (Section 7.3.2), we see that it is not possible to draw such a conclusion. The variables that determine artifact gesture frequency are much more complex than simply being affected by distance. In fact, one focus group participant suggested that in some senses distance had no impact:

> *"The fact that [R] was at home made no difference, we were all around the Smartboard, doing the same thing."*

Other participants suggested that being a remote participant was in some ways better than being collocated with the group, allowing them to perform parallel tasks (look up topics of discussion during the meeting) and adjust the view to meet personal needs (make the text large due to poor sight).

Finding 8 Remote researchers do not naturally interact with artifacts when they are not familiar with the technology being used, but adapt quickly when communication breaks down.

There were several instances in M3 where the interactivity and dynamics of the meeting were affected by distance. M3 was one of the early meetings where the research group did not have extensive experience with using shared applications like VNC.

Initially, the group put the spreadsheet up on the screen and simply talked about the data with no interaction (Minute 32 – 40). The remote user (R) then asked one of the local users to scroll the spreadsheet (Minute 41). Shortly afterwards, R realized that he could interact with the artifacts, and it is at this stage that mouse based artifact gestures events began to occur. R quickly became adept at using the mouse as a gestural based pointing mechanism. At one point in the meeting, R stopped referring to artifacts by pointing, causing some confusion among the local participants. Only after one of the local participants asked for clarification did R start using mouse-based gestures to point at artifacts again. This appeared to resolve the confusion of the local participants. Again, this sentiment was reflected in the focus group:

> *"Because our interactions always has as a focus either a document, looking at, commenting on, creating, and that document is on the smart board the whole texture of the meeting is incredibly sensitive to how well the smart board technology and the document manipulation works. Any glitches in there send things off the rails so quickly. We just lose momentum, which is a disaster."*

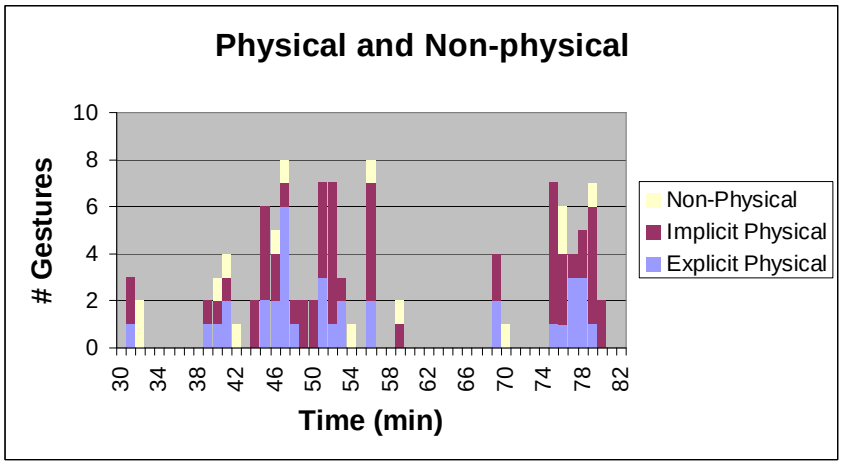


Figure 30: Physical and non-physical interaction in Meeting 11

Finding 9 The visual sensory streams used in M3 and M11 did not communicate artifact gestural events (physical artifact gestural events in particular) effectively to remote participants due to the lack of an appropriate visual sensory stream.

One of the key issues in distributed artifact-centric collaboration is how artifact gestures are communicated. In the context of the remote meetings observed (M3 and M11), artifact gestures were either communicated using the mouse to point at or select an artifact, touching the Smartboard (which acts like a mouse), or using a Smartboard pen to

annotate an artifact (circle or underline an artifact). Neither of these meetings had a visual reference space where the remote users could see a physical artifact gesture (pointing with the hand and not touching the Smartboard) being made in the context of the artifact. Physical artifact gestures were therefore not visible to remote participants.

In Figure 30, we show the same artifact gestures for M11 as shown in Figure 27, but display the artifact interactions based on whether or not they are explicit physical, implicit physical, or non-physical (using the mouse or Smartboard) artifact gestures. As can be seen, almost all artifact interactions in this meeting are physical gestures (made with the body) rather than mouse or Smartboard interactions. Since there is no visual stream that provides a reference space for this meeting, none of these artifact gestures are communicated to the two remote users. Thus, in M11 the remote participants were unable to see a large percentage of the artifact gesture events that occurred in this meeting. Note that the local presenter did not realize that these gestures were not being transmitted nor did the remote participants ask the presenter to use the technology to clarify which artifacts were being discussed (as they did in M3).

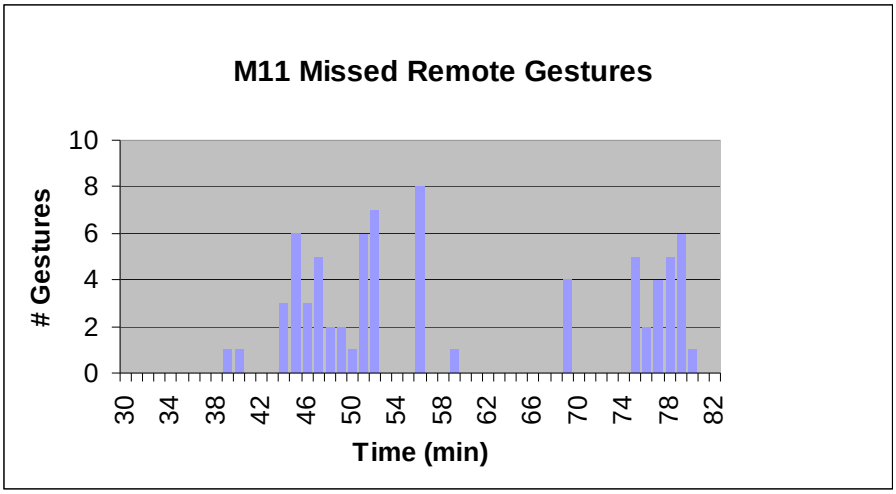


Figure 31: Missed gestures of major severity during Meeting 11

The severity of this problem can be seen from Figure 31. We apply a subjective scale to the severity (major or minor) of each missed gesture (physical gestures) made during M11. A missed gesture is assigned a "major" severity if the gesture is important to the understanding of the utterance ("this term" while pointing at a specific term in an equation) or the utterance is important to the conversation ("if the value goes below zero" while pointing to where a graph crosses the axis). If a gesture is not important to

understanding the utterance ("the first term in equation six" with a redundant pointing gesture to the first term in equation six) or it is not important to understanding the conversation (a comment like "this paper is very long" while pointing at an artifact as a representation of the entire paper) then the gesture is assigned a "minor" severity. Using this definition, each major missed gesture implies that the remote participants are missing a key detail of the conversation. As can be seen from Figure 31, this happens quite often with a total of 73 missed gestures of major severity and 74 missed gestures of minor severity (not shown).

Finding 10 Researchers often forget that there artifact interactions can not be seen by remote participants, and therefore do not utilize the technologies as effectively as possible to communicate artifact gestures.

It is worth pointing out that in M3 there were almost no missed gestures of a "major" severity. Recall that during this meeting, the remote participant (R) was the leader of the meeting. At the beginning of the meeting, R verbally referred to artifacts on the screen but did not perform any artifact gestures. Eventually, R was asked to point at an artifact to clarify a point. At this time, R realized that the mouse could be used as a pointing mechanism, and started to use the mouse relatively fluidly for this purpose (Minute 41-49 in Figure 25). At Minute 49, R started to describe a computational simulation (in another document), referring to artifacts with utterances but NOT making any artifact gestures. During this period there was some confusion. Not until Minute 68, when one of the local participants asked R to disambiguate "which column" R was referring to, did R start to use mouse based gestural interactions once again (Minute 68 – 72).

This problem is exacerbated when the presenter has a local audience as well as a remote audience (as in M11). When a presenter has a remote audience only, it is fairly clear that to the presenter that there is no one to "see" physical gestures. As we see above, presenters sometimes forget that they can use the technology to make artifact gestures. When a local audience exists and a presenter uses a physical artifact gesture, they know that the gesture is being communicated effectively to at least some of the participants. This is a familiar communication environment, and it is therefore relatively easy for the presenter to revert to physical artifact gestures rather than technology based gestures which would be visible to all participants. This is what we see in M11. Although the presenter sometimes used the Smartboard to highlight a feature on the screen, manipulate

artifacts (scroll a spreadsheet), or underline an equation with a Smartboard pen, a high percentage of the artifact gestures used in this meeting were physical based gestures (and "major severity" gestures in Figure 31). This is despite the fact that the remote participants are active in the discussion (that is, they are not just quiet participants).

Finding 11 Traditional distributed collaboration technologies (basic audio and video technologies) are awkward and difficult to use in the "real world".

Finally, it is worth pointing out that there were often issues in getting the basic audio and video collaboration technologies to function as desired. These problems ranged from struggling for 30 minutes to connect a remote user from Europe (see Appendix 15.5.1 for details), through users being disconnected part way through a meeting (and having to reconnect), to minor audio issues (such as the microphone of a remote user picking up the remote users typing and thereby disrupting the meeting). Since our focus is on gestural interaction, we do not analyse these faults in detail. With that said, such problems show up prominently in our coding and analysis. We would be remiss if we did not mention the wide variety of "basic" issues we observed with using "traditional" remote collaboration technologies on a day-to-day basis in the "real world".

7.3.4 Individual differences

In order to answer our research question "*How do researchers use advanced collaboration technologies?*", it is necessary to consider the individual differences between participants and how they utilize artifact gesture.

Finding 12 Different researchers use technology to communicate artifact gestures in different ways.

Some participants are extremely comfortable working with the Smartboard and are highly adept at manipulating artifacts (scrolling, changing pages, and opening documents) through the touch screen interface. These participants tend to use both physical and Smartboard based gestures interchangeably. Other participants are less comfortable using the Smartboard for manipulating artifacts, but are quite comfortable with the physical nature of working in front of a Smartboard. Such participants frequently use physical gestures, and might occasionally use the Smartboard to highlight an artifact using an underlining or circling type gesture. Other participants are more comfortable using the mouse as the gesturing mechanism, and typically stay seated at the meeting

room table rather than getting up and physically interacting with the Smartboard. Finally, some participants rarely interact with the technology at all during meetings (although this does not mean they do not contribute to the meetings in other ways).

The best example of this is the two presenters in M11. Both presenters were in the same physical environment. The first presenter did not use any artifact gestures at all and in fact did not approach the Smartboard (Minute 1 – 13, note that this time is not displayed in Figure 27 because no artifact gestures occurred). The second presenter made extensive use of physical gestures and occasionally used the Smartboard to interact with artifacts (Minute 30 – 80 in Figure 27). We see similar results in other meetings we observed, with individual researchers typically having a preferred "style" of interaction that they prefer.

7.3.5 Learning and adapting over time

To better understand our research question ""*How do researchers use advanced collaboration technologies?*" we also need to consider how researchers adapt to the technological environment in which they work. The exploration of adaptation also helps to answer two other important research questions: "*How do collaboration patterns change in the presence of technology?*" and "*How well do those technologies work*?" One of the key outcomes that emerged from this study is that participants are effective at learning mechanisms to overcome technological issues that obstruct the group from accomplishing its task.

Finding 13 Researchers adapt to the use of tools for collaborating around digital artifacts. This adaptation occurs both when the collaborators are collocated and distributed.

> *"At the beginning when we started it was really hard for me, people just talked, now we can not live without these things [Smartboards]"*
> *Focus Group Participant*

The ability to adapt was noticed in the CoTable case study (Chapter 5) and has been reinforced by the observations from our ethnography. This learning and adapting process can be partially seen in the way users utilized the Smartboard. Many users who had not previously been exposed to technologies like the Smartboard rapidly became comfortable using it for presentations and for brainstorming (sketching concepts and ideas). For example, in one of the later collocated meetings analyzed as part of this study (not M3,

M4, or M11) four participants interacted with the Smartboard. During this meeting participants interacted with artifacts through either physical (one participant) or Smartboard and physical (three participants) artifact gestural interactions. By the end of our observational period, most (but not all) study participants were comfortable performing artifact interactions using the Smartboard and several participants were highly adept at this skill. One focus group participant summarized the impact of the technologically sophisticated meeting rooms as follows:

> *"I think that you have to remember that this is a new project for all of us, I don't think any one of us has participated in anything like this, it started in this room with the Smartboards, we all came from different disciplines, a [DISCIPLINE X], a [DISCIPLINE Y], a [DISCIPLINE Z], I've never even heard of anything like this before, so I don't know how it would have been without [the technology]..."*

Finding 14 Researchers adapt technologies to perform complex interactions with digital artifacts in innovative and surprising ways.

The level of sophistication of how participants made use of the technology also changed over the duration of the observational period. Gesture usage in later distributed meetings became quite fluid, with participants passing control between local and remote users quickly and easily. In one instance, a remote user and a local user were showing a second local user some modelling data in a spreadsheet (shared using VNC). The two users were interacting with the spreadsheet almost simultaneously, with one user scrolling the document and the other user highlighting cells in the spreadsheet using the keyboard. One user was effectively the "document manipulator" while the other was the "data manipulator." The physical affordances of having a local user, a remote user, and simultaneous interaction facilitated this complex artifact manipulation. One of the participants from the focus group described the activity that led up to this interaction as follows:

> *"I started using the cursor, [the remote user] used the mouse, and it was brilliant."*

7.3.6 Physicality, engagement, and gesture

The need to naturally support physical gestures as part of an artifact-centric collaboration are exacerbated by the convergence of physical interaction (as supported by

devices such as the Smartboard), group interaction (as manifested by the increase in interaction we see when problem solving occurs), and artifact-centric gesture (as required by artifact-centric collaboration). Our observations help to answer our research question: "*How do collaboration patterns change in the presence of technology?"* In particular, our observations of artifact interaction help us to determine "*What communication channels are used to encode information during artifact-centric collaboration?"*

We hypothesize that touch sensitive screens, and the physical interaction that accompanies them, facilitates and encourages physical interaction both among users and with digital artifacts. This hypothesis, as yet untested, has been generated based on the findings in this study. We base this hypothesis on several factors, including the frequency of physical gestures, the frequency of multi-person interaction, and the frequency of turn-taking that we observed when participants are physically using the Smartboard. The use of a shared personal and task space (reference space) appears to encourage user and artifact interaction. It is exactly this shared space that our collaboration tools fail to support effectively.

7.3.6.1 Affordances of collocated interaction

We first discuss the affordances provided by the physical display and interaction device being within the same physical space as the user, considering the impact of these affordances on user and artifact interaction.

Hypothesis 1: Artifact gesture events are more frequent when users are physically close to the display and interaction environment.

Our observations indicate that users make artifact gestures often when collocated with the physical display. Our observations from M11 suggest that physical co-location with the display encourages physical gestural interaction. During M11, the speaker who presented the second paper (starting at Minute 39 in Figure 27) was physically close to the Smartboard. The speaker manipulated the artifacts using direct interaction on the Smartboard (Figure 32) and there were extensive physical and Smartboard gestures (Figure 30). For the first paper presentation (Minute 1 – 12 in M11), the presenter was far from the Smartboard and used no gestures (not shown in the figures because no gestures were made). An analysis of M4 reveals similar results. Almost all of the forty-two physical artifact gestures made in M4 are made collocated with the display. The majority

of these gestures are made from Minute 48 – 61 (see Figure 34) when participants L1 and L3 are both interacting with artifacts at the Smartboard at the same time (see Figure 35).

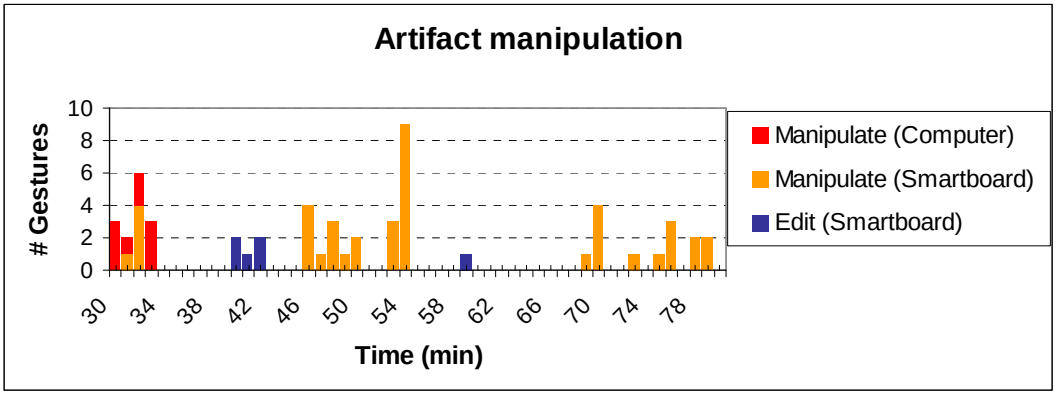


Figure 32: Artifact manipulation using the computer or Smartboard in Meeting 11

Increased distance from artifacts has been shown to reduce the amount of pointing with deixis, with deixis gestures the most frequent at arms length from the referent artifact [Ban04, Section 2.2.4.4]. Our research adds evidence in support of these results. Our hypothesis, in this case, suggests that gestural interaction with artifacts will be higher in the physically collocated space immediately in front of a Smartboard than in interaction spaces where the presenter is isolated from the presentation display. Interestingly, our study participants appear to feel that this collocation is not dramatically impacted by distance. For example, during the focus group, one of the participants stated

> *"… had we been all in the same room, we would have been sitting around a computer screen, doing exactly the same thing. The fact that [R] was at home made no difference, we were all around the Smartboard, doing the same thing."*

It should be noted that although our analysis suggests such a hypothesis, there may be a number of other factors that may influence artifact gesture event frequency. In fact, our analysis suggests that individual differences may also be a contributing factor (Section 7.3.4). Clearly, more research in this area needs to be performed.

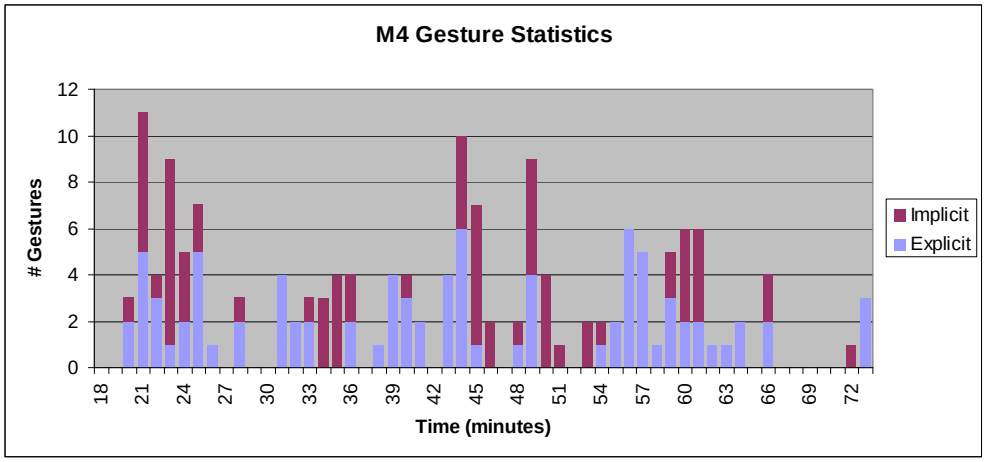


Figure 33: Meeting 4 gesture statistics

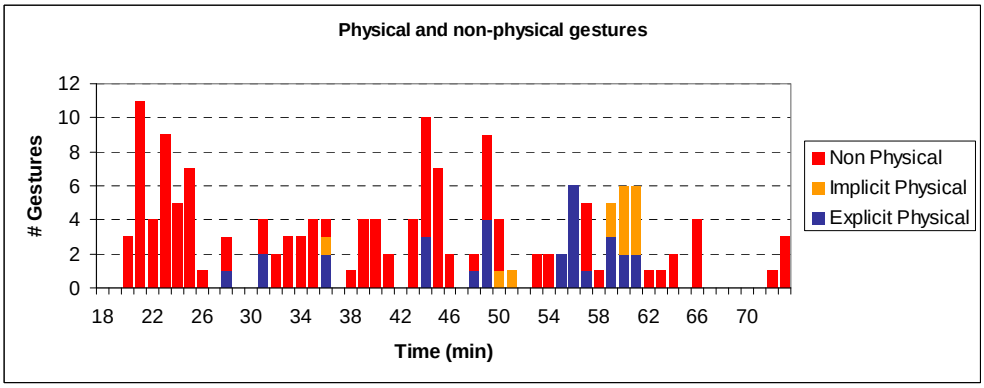


Figure 34: Physical and non-physical gestures in Meeting 4

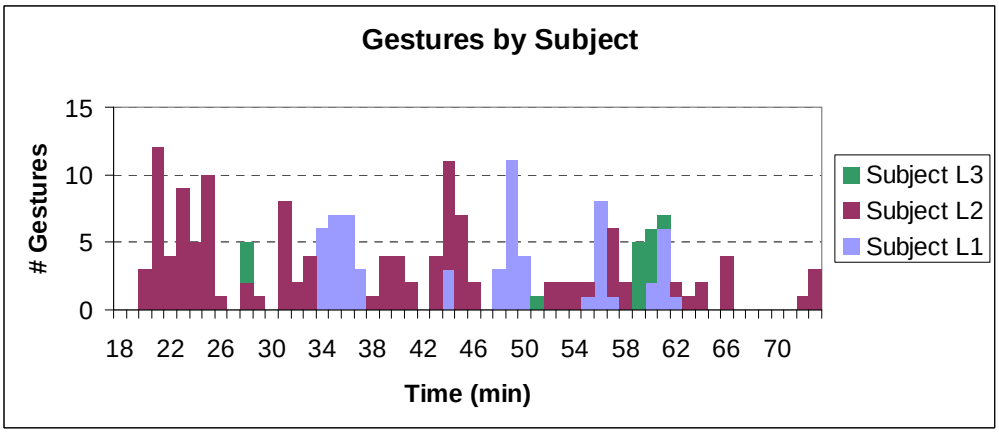


Figure 35: Gestures by subject for Meeting 4

7.3.6.2 Affordances of collocated engagement

We now consider how engagement and interaction between users might be affected by the collocation of the display, artifact interaction, and gesture spaces. We consider M4 in this regard. As discussed in Section 7.2.3.2 and Appendix 15.5.2, as M4 progressed it

became highly interactive. It slowly transitions from a description of a complex model, through an investigation of an odd feature of the model, to the discovery of a major problem with the model, and eventually to a solution to the problem. We hypothesize that:

Hypothesis 2: Task coupling, engagement, and interaction in artifact-centric collaboration are facilitated by the physical collocation of user space, work space, and gesture space.

Hypothesis 3: Touch sensitive devices that provide the collocation of user space, work space, and gesture space can facilitate artifact centric-collaboration.

Participants in the focus group alluded to the fact that the Smartboards and their ability to allow people to mark up documents (circle, underline, etc) without affecting the original document lowered inhibitions about interacting with artifacts on the screen:

> *"I think that if someone is giving a presentation, traditionally you are looking at overhead transparencies. Or you are looking at them writing on the blackboard, in groups larger than two or three people, your not very likely to get someone jump up and scribble on the blackboard on someone's else notes. Somehow because it's their notes, you aren't going to scribble all over them. But if you are looking at a paper on the Smartboard, that's fair game, you are totally uninhibited"*

In the three figures above we present the gesture statistics for M4 in three ways, presenting implicit/explicit artifact gesture events (Figure 33), physical and non-physical artifact gesture events (Figure 34[5]), and artifact gesture events as generated by different individuals (Figure 35). For the first 34 minutes of the meeting, almost all gestural interaction was performed with the computer mouse and by a single user (no physical co-location with the screen). There are no physical artifact gestures and there are almost no temporal artifact gesture interactions between individuals (multiple individuals alternately gesturing at artifacts). The artifact gesture events during the phases of the meeting from Minute 34 – 48, although being performed by multiple people, are primarily non-physical artifact gestures and there are almost no multi-individual artifact gesture exchanges. The exception to this is at Minute 44 where one of the participants (L1) gets up and physically goes to the screen to point at an artifact of interest.

[5] Note that Figure 34 shows emphasis gestures as well as explicit and implicit artifact events, so this graph displays slightly more gestural interactions than the other two figures.

Around Minute 48, the meeting starts to change in its nature. From Minute 48-51, L1 uses the Smartboard to solve a mathematical equation. Again, physical interaction with artifacts is prominent as is the direct manipulation of artifacts on the screen. The solving of this equation leads to an important insight into a discrepancy between the output of the computer model and the original data.

Minute 53-55 is a mouse-based exploration phase of the mathematical model by L2, with one physical artifact gesture by L1 (collocated with the display) and one by L2 (from the table). Minute 56 consists of a quick overview (given by L1) of the equation for one of the project members using the Smartboard, resulting in a number of physical artifact gestures. From Minute 57-58, L2 explores the model using mouse based artifact gestures. During Minute 59-62, a third subject (L3) gets up and approaches the Smartboard, asking L2 to look at a chart. At this point, most of the group is actively engaged in the discovery process and three of the participants (L1, L2, and L3) are actively pointing out artifact details in rapid succession. This is the peak of physical and personal interaction during M4. From Minute 62 – 66, L1 explores the model in more detail (using mouse based pointing gestures) and eventually finds the problem in the model (at Minute 64). There are no physical artifact gestures during this phase.

The above detailed analysis of M4 reveals that almost all of the forty-two physical artifact gestures are made collocated with the display. In particular, the dynamic problem solving part of the meeting, from the time where a problem was identified (Minute 48) to the time where L1, L2, and L3 finished problem solving (Minute 61), is the part of meeting when most of the physical artifact gestures in M4 occur (see Figure 34). During this period there are 33 physical artifact gestures and 18 non-physical artifact gesture (with the computer mouse by L2). Contrast that to the non-problem solving part of the meeting (Minute 0 – 47), where there are only 9 physical gestures and 88 non-physical gestures. Clearly, the interactive nature of the problem solving task resulted in a much higher frequency of physical artifact interaction events.

This time period is also when the coupling between group members is the highest. The meeting changes phases 22 times (see Section 7.3.1 for more details on phase changes in meetings), switching from description to discussion and back again. Figure 35 shows us that this time period also results in the most dynamic exchange between meeting

participants, with two participants interacting with the Smartboard and one participant interacting with the computer mouse over a four minute period (from Minute 59 to Minute 62). Thus not only is the phase changing rapidly, but so is who is interacting with the digital artifacts. It is worth noting that this increase in interactivity is also noticeable when watching the video recordings of the meeting. An individual's engagement level progresses from sitting sedately in his/her chair at the beginning of the meeting to several people standing in front of the Smartboard interacting with each other and the relevant artifacts in a very dynamic manner.

Our hypothesis that touch sensitive screens facilitate and encourage physical interaction both among users and with digital artifacts, is clearly worth further investigation. Our observations show that physical artifact interaction is very prominent when presenters are physically close to the display, but does not appear to be used as much when participants are far away from the screen. This finding is similar to that observed by Bangerter [Ban04] where pointing decreases based on distance from the referent.

In addition, physical co-location in front of an artifact display seems to facilitate and/or encourage engagement and interaction. In our observations, we see participants getting up from the meeting room table and engaging in artifact gestures and collaborator interaction in the physically collocated environment. From a CoGScience perspective, for **tasks** where the **nature of the material** is artifact centric, collocated displays and interaction appears to facilitate better **coupling** and promote **exploration** and **creativity**. At the same time, such an environment appears to support tasks that are **difficult**, **complex**, **urgent**, **competitive**, and **emotional**. In some meetings (M4 in particular), we see a convergence of physical interaction with artifacts at the same time as we see an increase in the interaction between collaborators. This suggests that the collocated physical environment provided by a touch screen interaction environment may be able to both facilitate and encourage collaboration.

7.4 Discussion

Our observations and analysis of artifact-centric, scientific collaboration have led to two primary types of outcomes. We have a generated a set of findings that help to answer

our research questions. In addition, as an exploratory study, our observations have also resulted in a set of hypotheses that require further study.

7.4.1 Findings

Our observations have provided us with the following findings. We consider these findings in the context of our research questions:

Objective 1: Develop a broad understanding of how scientific researchers collaborate.

- *How do collaboration patterns change in the presence of technology?*

Our results show that research meetings proceed in phases, that each phase can have a different goal or **task**, and that researchers switch between phases rapidly (**Finding 1**). Our observations also indicate that researchers adapt rapidly to the use of collaboration technology, in particular when that collaboration involves interaction with digital artifacts (**Finding 8**, **Finding 13**). They also show that adaptation can happen in novel and surprising ways (**Finding 14**). With that said, this adaptation is not consistent across users, and individual differences have a significant impact on adaptation (**Finding 12**).

During the focus group, participants indicated that their use of the technology had changed dramatically. Recall that although the research group had been using the room for several months before observations started, exposure to the advanced technology in the room was new to most people. One focus group participant stated:

> *"At the beginning when we started it was really hard for me, people just talked, now we can not live without these things [Smartboards]."*

Objective 2: Develop a deep understanding of how scientific researchers interact with digital artifacts when they collaborate.

- *What role do digital artifacts play in scientific collaboration?*
- *What information is lost when such collaboration takes place at a distance?*
- *What communication channels are used to encode information during artifact-centric collaboration?*

Our study demonstrates that artifact-centric collaboration is important to this research group. During the focus group, this was made clear by several statements about the importance of the group's use of desktop sharing (VNC) and artifact gesture interaction.

[P1] "The only down side as [P2] pointed out, was facial expression, the video was too small. Everything else was brilliant."
[Facilitator] "So what made it brilliant?"
[P1] "VNC, the fact that he could control the document, and I could control the document."

"The fact that [R] was at home made no difference, we were all around the Smartboard, doing the same thing."

"… what was really essential was for me to be able to highlight a part of the screen and say this is the cell I am talking about right now. I can do that remotely and we can both see exactly which part I was talking about."

"I HAD to have mouse control there, if I did not have mouse control, or at least the ability to point is what I am talking about when I mean control, it wouldn't work…"

Our observations show clearly that artifact gestures are used frequently, both in collocated and distributed meetings (**Finding 2**), but that such gestures are not used in all phases of a meeting (**Finding 3**). Artifact gestures frequency is often high during **loosely coupled description** tasks (**Finding 4**). Artifact gesture frequency is also often high during **tightly coupled discussion** tasks, in particular where there is **exploration**, **conflict**, **competitiveness**, and **urgency** (**Finding 5**). In some cases, a single researcher performs all of the artifact gestural interactions over a fixed period of time (**Finding 6**) while at other times many researchers perform artifact gestural interaction within a short period of time (**Finding 7**).

Objective 3: Evaluate advanced collaboration modalities and technologies for scientific collaboration.

- *How do researchers use advanced collaboration technologies?*
- *How well do those technologies work?*

Researchers in our studies made extensive use of advanced collaboration technologies over the duration of our study, including Smartboard interaction and sharing complex documents with remote participants. In our focus group the researchers indicated that the sharing of information (artifacts) was critically important to their research (see above). In general, researchers adapted quickly to using the technologies (**Finding 8**, **Finding 13**) and on occasion came up with sophisticated and innovative artifact-centric collaboration techniques that met their task needs (**Finding 14**). At the same time, there were some

significant problems with the technologies used during the research group's collaborations. The extensive use of artifact gesture, and in particular physical artifact gesture, were not effectively communicated to remote collaborators due to the lack of a mechanism that captured physical artifact interaction (**Finding 9**). Researchers either didn't realize and/or forgot that remote participants could not see their physical gestures, and therefore did not utilize the technologies to communicate gestures effectively (**Finding 10**). This suggests that remote collaboration tools need to be much more effective at either enabling natural interaction with the technology (so technology based artifact interaction is visible) or enabling the ability to capture physical artifact gestures and communicate them effectively to remote participants. Last, but certainly not least, our observations suggest that non-artifact centric collaboration technologies are not particularly robust. Although not discussed in detail in this dissertation, there were many problems with using "traditional" audio and video collaboration technologies throughout the meetings we observed (**Finding 11**).

7.4.2 Threats to Validity

Validity in qualitative research is considered differently than validity in quantitative research. As stated by McGrath and Brindberg, and reflecting our pragmatic approach to science (see Section 3.1.3), "Validity is not a commodity that can be purchased with techniques… Rather, validity is like integrity, character, and quality, to be assessed relative to purpose and circumstance" [BM85]. Creswell suggests that validity is a strength of qualitative research, but unlike validity in quantitative research, it is concerned with gauging the accuracy of an account of an event from the standpoint of the researcher, the participants, or the readers of the account [Cre03]. In our discussion below, we utilize the categorization of validity suggested by Maxwell [Max02].

7.4.2.1 Threats to Descriptive Validity

Descriptive validity is concerned with factual accuracy of the account described in the research. For example, if we record that an individual made a specific utterance or performed a certain action (e.g. a gesture), is our description of that action accurate? The main threats to descriptive validity in this study are the accuracy with which events are recorded and the fact that a single coder, the author, coded all of the events. The main

threat to descriptive validity in our coding scheme is whether utterances are transcribed correctly as utterances about artifacts and whether those that are artifact utterances are deictic or not. This is critically important as utterances are used to define our higher level gestural communicative events. To mitigate this threat, the mechanism used to define our codes is relatively mechanical and is done during a post-hoc analysis of the video taped meetings that were observed. This helps us to decrease the threat of transcription errors on utterances, as the transcription can be performed carefully through analysis of the video.

Although we did not perform an inter-coder reliability test directly on the coding scheme used in this study, this coding scheme was also utilized in the gesture study described in Chapters 8 through 11. Section 8.3.1.2 presents how the gesture coding scheme was used in the gesture study, including an inter-coder reliability analysis (Appendix 15.9.1). This analysis demonstrates that the coding scheme used in this chapter is consistent at generating high-level communicative gesture events across multiple coders. This inter-coder reliability helps to mitigate the threat of personal biases in the coding of the coding performed in this study.

7.4.2.2 Threats to Interpretative Validity

Interpretative validity is concerned with the accuracy of the interpretation of the meaning that observed events and behaviours have to participants. Interpretative validity is specific to qualitative research, as it inherently suggests that meaning is constructed by the researcher on the basis of a combination of the observation of events and a participant's accounts of those events. Interpretative validity is different from descriptive validity in that while consensus can be reached about categories used in description (an utterance, an action), consensus about the categories used in interpretation rest on the participant's own view of the events. Maxwell uses the example of the utterance and action of a teacher yelling at a student in class. Although descriptively this event can be well defined, the interpretation of whether the teacher was really mad at the student or just trying to get control of the class relies on the perspective of the teacher.

The main threat to interpretive validity to this study is whether or not our high-level gestural constructs, such as emphatic and implicit gestures, capture the meaning that our participants are attempting to communicate. In addition, our analysis suggests that these

interactions are important constructs for communication, and we therefore need to address the issue that our interpretation of the importance of these actions may not be reflected by the participants.

We mitigate the threats to interpretive validity using a number of methods suggested by Creswell [Cre03]. First, we incorporate *triangulation*, through using both our observations as well as our focus group, to support our claim that artifact interaction is indeed important to the research group considered in our study. Second, we spend *prolonged time in the field*, developing an in-depth understanding of the activities of the research group. This helps the observer to develop a better understanding of how the research group works with artifacts in a wide variety of circumstances. Third, we balance this individual time in the field with a discussion of the potential *bias* that having a single observer might impart on the analysis and how we have attempted to mitigate against that bias (see Section 7.4.2.1). Fourth, we provide a *deep and rich narrative* (Appendix 15.5) that describes the meetings that we analyse as part of our study, allowing other researchers to make their own interpretations of our analysis.

7.4.3 Hypotheses

Our study of artifact-centric, scientific collaboration is exploratory in nature, with one of the goals to generate theory and hypotheses that suggest a new understanding of the phenomenon in question. The findings above lead us to the creation of a set of hypotheses about distributed, artifact-centric, scientific collaboration. Note that we do not perform any empirical hypothesis testing in this study, but instead are concerned with the generation of new hypotheses that inform our further study.

The first set of hypotheses is derived from our analysis of the dynamic nature of research meetings and the physical nature of the interaction technologies our participants were using. Our analysis suggests that the affordances of a physical reference space that collocates task space (the artifacts) with personal space (the physical space a researcher occupies) has an impact on the frequency and types of gestures that are performed. In particular, we hypothesize that:

- ***Artifact gesture events are more frequent when users are physically close to the display and interaction environment.***

- ***Task coupling, engagement, and interaction in artifact-centric collaboration are facilitated by the physical collocation of user space, work space, and gesture space.***
- ***Touch sensitive devices that provide the collocation of user space, work space, and gesture space can facilitate artifact centric-collaboration.***

Our fundamental analysis in this study focuses on the frequency of artifact gestures and the impact that distant collaboration technologies have on those gestures. Our analysis of the use of gesture shows that artifact gestures are utilized extensively to refer to digital artifacts as a part of research meetings. Our analysis also shows that even state-of-the-art collaboration infrastructure does not support the transmission of artifact gestures effectively. Given that we use gesture naturally in everyday communication and our analysis shows that this gesture use often translates into artifact-centric gesture when discussing complex scientific data, it seems obvious that we should build better tools to support artifact-centric, scientific collaboration. Such a statement assumes that our inclination to use gesture serves some purpose, and that researchers make use of the artifact-centric gestural communication that we observed in our study. Our current study does not provide any evidence to support this premise. The frequent use of artifact gestures implies that these gestures communicate information.

Our final set of hypotheses explores whether or not artifact gestures are observed, decoded, and processed by researchers. This leads us to the following hypotheses:

- ***Researchers will attend to artifacts when they are used as part of a presentation.***
- ***Researchers will attend to artifacts more frequently when gesture is used to draw attention to an artifact.***
- ***Researchers will have a better understanding about artifacts, how they are used, and the information they contain when gesture is used to refer to those artifacts during a presentation.***

It is these three hypotheses that we chose to explore in more detail in Chapter 8 through Chapter 11. The hypotheses involving the collocation of task space and user space are left for future research.

8 Understanding the Use of Gesture – An Experiment

The gesture study presented in the following chapters is targeted at meeting one of our key research objectives:

- *Develop a deep understanding of how scientific researchers interact with digital artifacts when they collaborate.*

In particular, it focuses on exploring in depth one of the research questions under this objective, that of:

- *What communication channels are used to decode information during artifact-centric collaboration?*

This in turn helps us to answer one of the key research questions we need to answer in order to deliver effective and reliable collaboration tools for distributed, artifact-centric, scientific research:

- *What human communication channels need to be supported for artifact-centric collaboration?*

In the analysis of collaborative scientific research presented in Chapter 6 and Chapter 7, we see that such collaborative research occurs in a wide range of scenarios, ranging from formal presentations to large audiences to intimate research discussions among a small research group. We also show that gesture is an important element of the process that researchers use to communicate scientific concepts. This is particularly the case when technologies such as touch sensitive devices (such as a Smartboard [Smart]) facilitate this class of interaction through direct interaction with the artifacts. Chapter 6 and Chapter 7 also point out some of the flaws with current, state-of-the-art collaboration technologies. In particular, Chapter 7 demonstrates that while gesture is used extensively when researchers are discussing digital artifacts, current collaboration technologies do not effectively transmit this information to remote collaborators.

It is important to note that the studies presented previously do not deal with the decoding, or receiving side of the communication process. In Chapter 7, we show that individuals use gesture to communicate information, but how do we know what impact gesture has on the decoding and understanding process of the receiver. Although the communicative aspects of gesture have been studied extensively (see Section 2.2.4 for a

detailed discussion), the decoding process (the process of understanding the communicative intent of the gesture) has received much less study than the encoding process (the process of encoding meaning using gestures) [BC06][BC00]. As recently as 2006, Bavelas and Chovil point this out as an important gap in the gesture literature, stating that "*…there are still far fewer decoding than encoding studies*" [BC06].

In order to help inform the design process for the creation of new collaboration tools, it is essential to understand how scientific collaborators process and decode the meaning and intent that is communicated through artifact gestures. It is not enough to know that researchers use artifact gestures extensively. Nor is it enough to know that current collaboration tools do not communicate that gesture effectively. These two facts cause a problem if, and only if, artifact gesture is important to the effective understanding of the intended communication. We do not need to design collaboration tools that transmit gestural interaction if gestural communication does not add to the efficacy of the communicated message. The experiment presented in the following chapters studies the decoding process utilized by researchers when viewing a scientific presentation.

The presentation of this study is spread across several chapters. This chapter (Chapter 8) presents the study design. Chapter 9 analyzes the aspects of the study that are not affected by our experimental interventions. Chapter 10 presents our statistical analysis of our experimental intervention. Finally, Chapter 11 casts the results from Chapter 9 and Chapter 10 in the context of our research objectives, research questions, and study hypotheses.

8.1 Situation

The collaboration scenario studied in this experiment is that of a remote presentation involving a complex research problem. In a research environment, presentations are widely utilized to disseminate research results. Such presentations can take many forms. It is therefore necessary to elaborate on the type of research presentation considered in this experiment. The presentation being studied is a presentation about global warming, its causes, and its effects. The goal of the presenter is to convince the audience that humans need to take immediate action in order to mitigate the impacts of global warming on our society.

This study focuses on several key aspects of the information decoding process:

1. Do participants attend to (look at) artifacts when watching the presentation?
2. Do gestures assist in drawing attention to those artifacts?
3. Does gesture increase the understanding of the topic being presented?

A three-way communication process is simulated in this study. The presenter is trying to convince the audience of a specific scientific argument. A questioning audience member asks questions of the presenter. We refer to this individual as the Devil's Advocate, partly because he challenges the presenter and partly because the he is actually dressed like a devil (he wears a hat with horns and uses fire and smoke as props). Note that the Devil's Advocate is actually the same person as the presenter, playing two separate roles in the video. The study participants are passive observers of the presentation and the dialogue between the presenter and the Devil's Advocate.

The interaction between the presenter and Devil's Advocate is a recorded interaction, and is played back to subjects in the study. During the presentation, the presenter makes extensive use of a whiteboard to create and manipulate artifacts in the support of his argument. It is this interaction with the artifacts that is the focus of this study. From a distributed collaboration perspective, it is as if the subject is viewing the presentation, including the dialogue between the presenter and the Devil's Advocate, using a web streaming or non-interactive video conferencing technology.

8.1.1 Exploring the collaboration task using the CoGScience Framework

Task	Function	Process	Channel
Execute - Performance Choose - Intellective - Decisions	Express ideas Engage audience Explain topic Make decisions	Conversation - Engagement - Trust Work Object - Create - Modify - Manipulate - Awareness - Monitor	Verbalize Paralinguistic Gesture - Manipulation - Kinetic/spatial - Pointing Facial expression Body language Gaze awareness Workspace awareness

Table 4: Global Warming presentation CoGScience task breakdown

The CoGScience Framework provides four levels at which to consider task: the **task** classification, the **functions** that need to be performed to accomplish the task, the

communication **processes** required to carry out those functions, and the communication **channels** that are required to provide that functionality. The goal of this part of the CoGScience Framework is to help us drill down from the high-level task to a set of concise human communication channels (see Table 4). Below we briefly consider each level of the CoGScience **task domain** in the context of the global warming presentation. This task analysis is considered in more detail in Appendix 15.6.

The main **task** of a presentation is to deliver (**execute**) the presentation. That is, it is a **performance** driven task. In the global warming presentation, although the main task is to execute (give the talk), the task also has elements of a **choosing** task. That is, the speaker is trying to convince the audience that they have a choice to make and that they should chose to take action. From the presenter's perspective, the task is **intellective** (making the correct choice), as the speaker is convinced that there is a specific, correct outcome. From the audience's perspective there may be no correct choice (this is why there is controversy on this topic after all) and therefore to some it may be a **decision** task (choosing an alternative).

The main CoGScience **functions** required to accomplish this task are to **express ideas, engage** the audience, and **explain** a complex topic**.** In addition, the goal is to help the audience **make a decision** about what action to take. In particular, we target this study at helping to increase our understanding of how gesture and facial expression affect the ability to **express ideas** and make **decisions**. In a normal distributed presentation, the ability to **discuss** would also be an important communication function to consider. In this study, we eliminate this function from consideration by controlling for it as part of the study.

Several **processes** are important to this type of communication task. Processes that support the **conversation** include **engagement** and developing **trust**. Because of the whiteboard use in the presentation, the **work object** is prominent in our analysis. Processes that support the **work object** include the ability to **create**, **modify**, and **manipulate** artifacts as part of the presentation. It is also necessary for the audience to be **aware** of the work space and to **monitor** how the speaker is interacting with that workspace. In fact, it is within this workspace that we perform our experimental

intervention by controlling for how much workspace awareness subjects in different conditions have (see Section 8.3 for details).

Finally, the CoGScience Framework also suggests that it is necessary to consider a range of human communication **channels** for this task. We briefly list these communication channels below, but the reader is referred to Appendix 15.6 for a more detailed analysis. From an aural perspective, the ability to use both a **verbal** and a **paralinguistic** channel is important. Since artifact interaction is the focus of this study, all forms of **gesture** (**manipulation**, **kinetic**, **spatial**, and **pointing**) are of high importance. **Facial expression**, **body language**, **gaze awareness**, and **workspace awareness** are also critically important. In fact, the main variables that we manipulate in this study (**gesture** visibility and **facial expression** visibility) are CoGScience communication channels. Our experiment uses a multi-factor design to consider the impacts of gesture and facial expression visibility on artifact attention and understanding (see Section 8.3.2 for a description of these experimental conditions).

8.2 Hypotheses

The ability to determine whether or not collaborators attend to gestural interaction when researchers are collaborating at a distance is critically important. If we want to provide design guidelines for building tools for remote, scientific collaboration we need to understand how gesture is used, how gesture is decoded, and how gesture impacts understanding in remote scientific collaboration. In Chapter 7, we show that gesture is used extensively in many types of research meetings, both collocated and distributed. Unfortunately, we do not know to what degree researchers use this gestural information as part of the decoding and understanding process. The first step in determining whether artifacts are important in understanding is to determine whether or not artifacts that are referred to during a presentation are attended to (looked at) by the observing researchers. This leads us to our first hypothesis:

Hypothesis 1: Researchers will attend to artifacts when they are used as part of a presentation.

This leads immediately to a second question. That is, when a gesture is made at an artifact does the pointing action effectively draw attention to that artifact? Thus, the second hypothesis we consider is:

Hypothesis 2: Researchers will attend to artifacts more frequently when gesture is used to draw attention to an artifact.

A third important question that we need to consider is, does gestural interaction help in the transfer of knowledge and increased understanding? That is, is there any evidence that artifact specific gestural interaction is useful in helping the decoding and understanding of the message being presented? In particular, we are interested in determining the impact of gesture on the participant's understanding of the structure of the artifacts used in the presentation, their understanding of the role the artifacts play in the presentation, and their understanding of the information content that is contained in the artifact. Thus, a third hypothesis is:

Hypothesis 3: Researchers will have a better understanding about artifacts, how they are used, and the information they contain when gesture is used to refer to those artifacts during a presentation.

In addition to trying to provide evidence about the efficacy of gestures in artifact-centric collaboration, this study also considers other factors that may impact distributed face-to-face communication. There has been a recent resurgence in the social psychology field in exploring the close ties between verbal and non-verbal communication in face-to-face communication. Much of this research explores how facial expression and gesture are used in combination with verbal communication (see Section 2.2.4 and [BG07][BV06] for more details). At the same time, the CSCW community has been exploring the use of visual information for communicating much more than just the "talking head" that one traditionally sees in a video conferencing environment [NSK+93]. In this study, we also explore the impact of communicating facial expression as part of distributed, collaborative research. In particular, we want to ask similar questions about facial expression to those we asked about gesture. Our fourth hypothesis is:

Hypothesis 4: Researchers will attend to facial expression when it is communicated as part of a presentation.

If facial expression is attended to during a presentation, such attention may cause a distraction from the attention paid to the artifacts being discussed in the presentation. If the artifact is a key tool to communicating knowledge in the presentation, drawing attention to the face of the speaker may remove attention from the artifacts in question. This leads us to our next hypothesis:

Hypothesis 5: Researchers will attend to the artifacts used in a presentation less when facial expression is visible as part of the presentation.

Assuming that researchers do not attend to artifacts as often, it follows that this lack of attention may also affect the communicative power of the artifacts. Thus we hypothesize that:

Hypothesis 6: Researchers will have a poorer understanding about artifacts, how they are used, and the information they contain when facial expression is visible as part of the presentation.

Given that there are two independent variables that we are considering in the hypotheses above (whether gesture is visible and whether facial expression is visible), it is important to ask whether the two independent variables interact with each other. That is, does the effect of communicating facial expression have any impact on the effectiveness of communicating with gesture and does communicating gesture have any affect on the effectiveness of communicating with facial expression. We hypothesize that facial expression will draw attention away from the artifacts, and therefore knowledge about the artifacts will be reduced when facial expression is communicated. In some sense, Hypothesis 6 captures this expectation, in that if we use a multi-factorial design, a decreased understanding will result across the gesture factor.

8.3 Treatment

In order to answer the questions presented above, we make use of a video that presents an argument around global warming. The video is part of an extensive series of videos on global warming, which recently resulted in the publication of a book on this topic [Cra09]. We control for two main independent variables in this study, the visibility of facial expressions and the visibility of gestural interaction. The study is therefore a factorial design, with two factors (Head and Gesture), and within each factor, we have two levels, visible and not visible. The study is a between subjects design where each subject sees only one of the treatments. Human research ethics approval for this study was obtained from the University of Victoria Human Research Ethics Board.

The video is ten minutes in length and switches between the main presenter and the Devil's Advocate who interacts with, and questions, the presenter[6]. Both the presenter

[6] Recall that the two characters in the presentation are actually the same person but are dressed differently and act as two separate characters during the presentation.

and the Devil's Advocate characters make extensive use of physical props, sometimes to make a point, sometimes for entertainment value. Although the video discusses a serious topic, the presentation is entertaining and somewhat tongue in cheek. The presenter is on screen and/or speaking for approximately 8 minutes of the video, while the Devil's Advocate is on screen and/or speaking for approximately 1.6 minutes. The presenter utilizes several key diagrams on a whiteboard to make his point. These whiteboards sessions, and the information presented on them, are the primary artifacts that we consider in this study.

The main diagram used as part of the presentation is a version of Pascal's Wager. Published as part of Blaise Pascal's 17th century defence of the Christian Religion (note 233 of Pascal's *Pensées*), Pascal's Wager is a technique for making a decision under uncertain conditions. According to this decision theory methodology, when faced with more than one action which could give rise to more than one outcome, a rational approach is to identify the possible outcomes, determine their values (positive or negative outcomes), and assign each a probability. The expected value of taking an action is then the value multiplied by the probability. Pascal's Wager suggests that the correct action to take is the one with the highest expected value. Pascal used this argument to explore how an individual should act faced with the improvable question as to whether or not God exists. In the video used for this study, the presenter explores the possible actions that society can take, given the uncertainty around whether or not humans cause global warming. An example of the diagram that represents the exploration of this question using Pascal's Wager is shown in Figure 36.

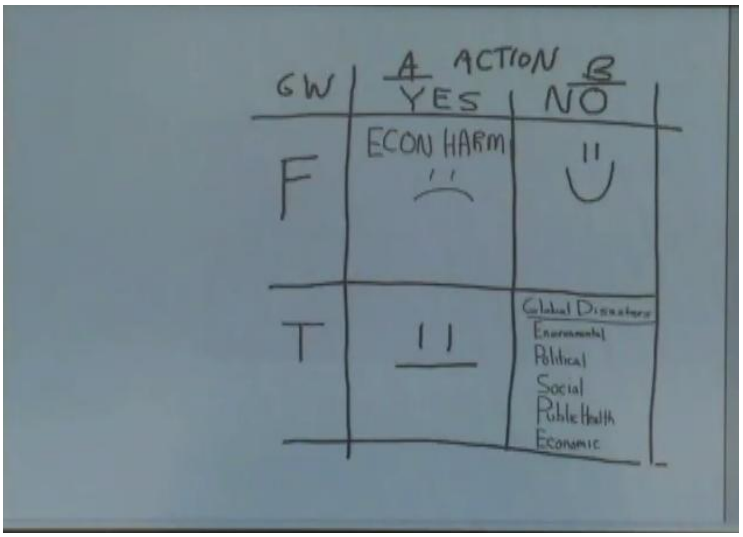


Figure 36: Pascal's Wager applied to whether or not humans cause global warming

8.3.1 Acts, Scenes, and Area of Interest

In order to create the four conditions required for this study, it is necessary to identify the sections of the video that have gestural interactions that refer to artifacts. We treat the video used in this study like a play or piece of theatre, breaking the movie down into acts and scenes. We further partition each scene into a set of mutually exclusive Areas of Interest (AOIs), each of which encompasses an area on the screen in which something "interesting" occurs. A summary of each act, the number of scenes per act, and the number of scenes per minute are outlined in Table 5. These concepts are explained further below.

Act	Description	Duration (s)	Scenes	Scene/min
Act 1	Introduction	64	12	11.25
Act 2	Whiteboard - Action or No Action	23	13	33.91
Act 3	Dialogue	28	10	21.43
Act 4	Whiteboard - Pascal's Wager	113	67	35.58
Act 5	Dialogue	113	17	9.03
Act 6	Whiteboard - Explaining Probabilities	30	20	40.00
Act 7	Dialogue	204	48	14.12
Act 8	Calibration	14	7	30.00
Whiteboard	All whiteboard acts	166	100	36.14
Total	Entire video	589	194	19.76

Table 5: Acts and Scenes

8.3.1.1 Acts

We divide the video into eight acts. The acts used here are the equivalent of the meeting phases encountered during the ethnographic study (see Sections 7.1.2 and Section 7.3.1). Each act of the video encompasses a phase of the collaboration session that utilizes a different style of presentation. Three of these acts (Acts 2, 4, and 6) take place in front of a whiteboard and contain gestural interactions with the artifacts displayed on the whiteboard. These acts are highly dynamic, with the presenter making extensive use of gesture. Four of the acts (Acts 1, 3, 5, and 7) consist of dialogue between the presenter and the Devil's Advocate (where the camera view is switched from one individual to the other as they speak). In these phases, the presenter's actions are less dynamic and there are fewer interactions with artifacts or props (although some props are used). The last act (Act 8) is a calibration act, which provides a mechanism to determine how well our measure of attention (eye fixation) is performing at the end of a given subject's participation. The acts vary in length, with an average act length of 82 seconds, with the longest being 204 seconds and the shortest 23 seconds. We exclude Act 8 in much of the discussion below (except when we discuss calibration) as it has no dialogue or interaction.

8.3.1.2 Scenes

We then divide the acts into one or more scenes, where a scene represents a contiguous related action by one of the individuals in the video. For example, one scene might consist of the presenter making a series of statements, where the next scene might be the Devil's Advocate challenging a point the presenter made. For analysis purposes, the entire video is divided into scenes, with no gaps between adjacent scenes. Scene divisions are done very accurately (measured in 10s of milliseconds), as a granularity of a second (and even 100 ms) was considered too course to provide the accuracy required. We test the validity of our scene decomposition with a scene inter-coder reliability analysis. A description of the inter-coder reliability analysis can be found in Appendix 15.9.1.

We base the granularity of the division of the whiteboard acts on the coding scheme that emerged from our ethnography in Chapter 7 (Section 7.1.2 and Section 7.1.3). The types of actions in the video that generate scene breaks are:

- The character displayed on the screen changing (the presenter to the Devil's Advocate or vice versa);
- The speaker moving within the scene (necessary to keep track of facial features);
- The speaker making a gesture (artifact or emphatic); and
- The speaker manipulating an artifact (either writing or erasing something on the whiteboard or manipulating a physical prop).

The number of scenes in a given act typically depends on the level of gestural interaction that occurs, as we need to provide a detailed analysis of scenes that have high levels of such interaction. For example, a monologue by the presenter, although it may contain many utterances, is considered a single long scene because there are no artifact interactions within it. In these cases, we would only have a new scene when the speaker changes. On the other hand, the acts that have artifact interaction in them (Acts 2, 4, and 6) are subdivided relatively finely to enable our analysis of that interaction, with scenes typically corresponding to single utterance and gestural phrases. Like the meetings analysed in our Ethnography (Chapter 7), we performed a full analysis of the presentation using the coding scheme from Section 7.1.2 and Section 7.1.3. During the presentation (and in particular during the whiteboard scenes) there are up to 22 artifact gesture events in a single minute. A summary of the number of explicit and implicit artifact events (these are defined in the next section) in each of the whiteboard acts is given in Figure 37.

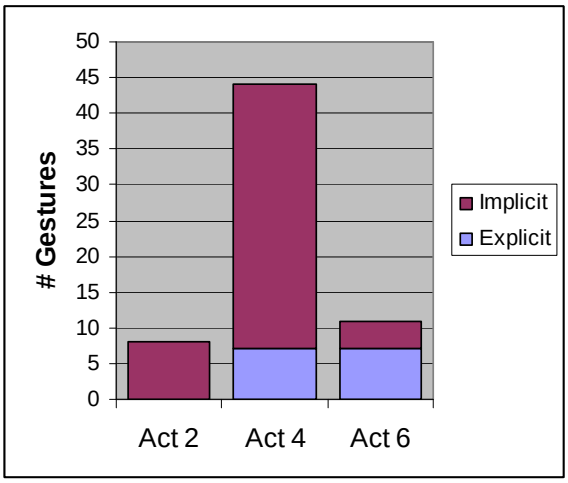


Figure 37: Explicit and implicit artifact communication events

As seen in Table 5, the number of scenes in an act varies dramatically. This is a result of both the varying durations of the acts (23 seconds for Act 2 to 204 seconds for Act 7)

and the granularity at which scenes are created based on the type of interaction. Considering the scenes/minute column in Table 5, it is clear that in the whiteboard acts (Act 2, 4, and 6) our analysis utilizes significantly more scenes per minute (over 30) than in the dialogue acts. On average, during the three whiteboard acts we utilize a new scene once every 1.66 seconds. Over the entire video, there are 194 scenes, with an average of 19.8 scenes per minute or one scene every three seconds.

8.3.1.3 Areas of Interest

In order to determine what subjects are looking at during the study, it is necessary to create a set of Areas of Interest (AOIs) within each scene. An AOI is a region of the screen that holds interest for us in our analysis during the duration of a scene. We want to be able to determine if a subject's gaze is focusing on such an area of interest. Typical AOIs would be the speaker's face, an artifact on the whiteboard, a physical prop that is being used for emphasis, or the speaker's hands during an emphatic gesture. In particular, because we are interested in artifact interaction, AOIs often revolve around artifacts that are being referred to as part of a gestural interaction. We test the validity of our AOIs in a similar way we did for our scenes, with an AOI inter-coder reliability analysis. A description of the AOI inter-coder reliability analysis can be found in Appendix 15.9.2.

An AOI is a shape, usually a rectangle, which surrounds the area of the screen in which we are interested. AOIs are divided into types, where the type of AOI provides us with information about what is contained within that AOI. We make use of five primary AOI types:

- FacialFeature AOI: The area of the scene where a person's face appears.
- EmphaticGesture AOI: The area of the scene where the speaker is using gesture for emphasis.
- ExplicitPointArtifact AOI: The area of the scene that encompasses an artifact that is referred to during an explicit communication event (e.g. using a deictic utterance such as *"this is the largest value in the diagram"* while pointing at the number 42 in a diagram on the whiteboard).
- ImplicitPointArtifact AOI: The area of the scene that encompasses an artifact that is referred to during an implicit communication event (e.g. using a non-deictic

utterance such as *"42 is the largest value in the diagram"* while pointing at the number 42 in a diagram on the whiteboard).

- ArtifactManip AOI: The area of the scene that encompasses an artifact as it is being physically manipulating by the speaker (writing, erasing a line, moving an artifact).

The four gesture-based artifact AOIs are derived from the compound communication events defined in our ethnography (see Section 7.1.3). We revisit these definitions here for clarity. Emphatic events are compound gesture/utterance events in which the gesture is used for emphasis only (there is no artifact gestural interaction). From a CoGScience Framework perspective, these are **kinetic**, **spatial**, or **rhythmic** gestures. EmphaticGesture AOIs encompass the area where the gesture is made.

Recall that our ethnography (Section 7.1.3) divides up the CoGScience artifact **pointing** gestures into two classes of communication events, implicit artifact events and explicit artifact events. We use this classification throughout this study as well. Implicit artifact events are those events that combine a pointing gesture with a non-deictic utterance that refers to an artifact on the screen (the phrase *"the number 42"* while pointing at the number 42 on the screen). We call this an implicit event because the referent artifact is implicit in the utterance. ImplicitPointArtifact AOIs surround the artifact that is referred to by the gesture component of an implicit communication event.

Explicit communication events combine a pointing gesture with a deictic utterance (the phrase *"… this is the number"* while pointing to the number 42 on the screen). We call this an explicit gesture because the artifact that is the referent of the communication cannot be implied from the utterance (as it can in an implicit communication event). That is, the utterance is meaningless without the explicit pointing gesture that refers to the relevant artifact. ExplicitPointArtifact AOIs surround the artifact that is referred to by the gesture component of an explicit communication event.

Finally, we consider CoGScience **manipulation** gestures. Manipulation gestures are gestural interactions that are used to manipulate artifacts, such as moving artifacts, creating an artifact through writing, or underlining an artifact for emphasis. An ArtifactManip AOI surrounds the artifact that is being manipulated. For a more detailed

discussion of the communication events defined in our ethnography, please refer to Section 7.1.3.

Note that the last three AOI types are all artifact-based AOIs. That is, they are AOIs that surround an artifact of interest in the presentation. We differentiate between the three types of artifact AOIs based on the type of gesture used to refer to the artifact, facilitating our analysis of the different type of artifact gestures. It is important to remember that these AOIs surround the artifact that is pointed at, not the pointing gesture itself. In our analysis, we are interested in determining whether gestures draw attention to artifacts, not whether gestures draw attention to gestures.

For all of the main classes of AOI, we also make use of "post-action AOIs". A post-action AOI is similar to the AOIs above, except that it occurs in the scene that immediately follows the scene in which the action actually took place. During our pre-study testing, it was noticed that subject gaze lingered on areas of the screen that were of interest, even though the speaker had gone on to performing another action. As a result, we created post-action AOIs. For example, in Act 2, Scene 9 (denoted Scene 2-9), the presenter performs an implicit artifact gesture. In that scene, we would have an ImplicitPointArtifact AOI. In Act 2, Scene 10 (Scene 2-10), the speaker is no longer making the pointing gesture and is in fact not pointing at anything. Since there is no dramatic action occurring in Scene 2-10, it is possible that the subject's gaze will still be focused on the artifact referred to in Scene 2-9. In order to capture this, and at the same time differentiate such focus of attention from that paid to the original pointing action, we utilize FacialFeaturePost, ExplicitPointArtifactPost, ImplicitPointArtifactPost, EmphaticGesturePost, and ArtifactManipPost AOIs.

For most of the scenes in the dialogue acts of the video (Acts 1, 3, 5, and 7) there are typically one or two AOIs. There is always a FacialFeature AOI, with other AOIs being utilized to capture interesting behaviour such as emphatic gestures, artifact gestures, and artifacts being manipulated. In the whiteboard acts (Acts 2, 4, and 6), there is always one, usually two, and sometimes three AOIs. There is always a FacialFeature AOI. For many scenes, there is one ExplicitPointArtifact, ImplicitPointArtifact, EmphaticGesture, or ArtifactManip AOI (primarily because scenes are defined by these gestures). Some scenes also have a post-action AOI.

8.3.2 Treatment Conditions

Our study controls for two factors, gesture visibility and facial feature visibility. In each of these factors, we have two levels, visible and not visible. We are primarily interested in analyzing the decoding process when gestural interaction is used with artifacts on the whiteboard. As discussed in Section 8.3.1.1, we have three acts in which the whiteboard is used. Two of these scenes (Scene 4 and 6) utilize a diagram representing Pascal's Wager (Figure 36) while the other whiteboard (Scene 2) is a diagram with text (Figure 38). All subjects see identical videos during the non-whiteboard scenes, with subjects in each condition seeing slightly different versions during the three whiteboard scenes. The four different conditions are described below.

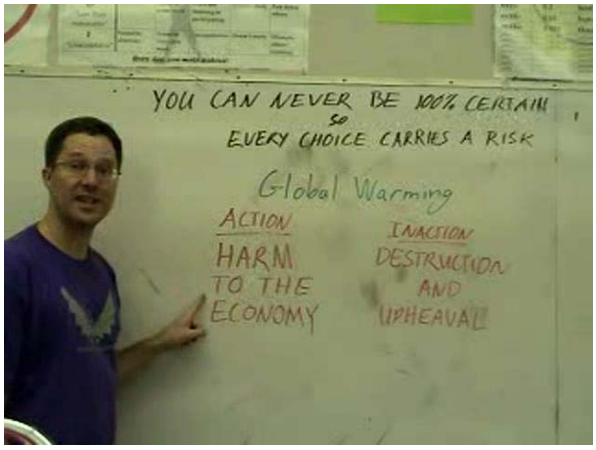


Figure 38: A whiteboard scene with an artifact gesture.

8.3.2.1 The Yes Gesture Yes Head Condition (YGYH)

The base case for the study is the original video, which represents face-to-face communication as closely as possible. During the whiteboard acts, the video shows the presenter standing in front of the whiteboard. It is therefore possible to see both gestural interaction and facial expression. We call this the Yes Gesture, Yes Head condition (YGYH). Images showing two scenes from this condition are given in Figure 38 and Figure 39.

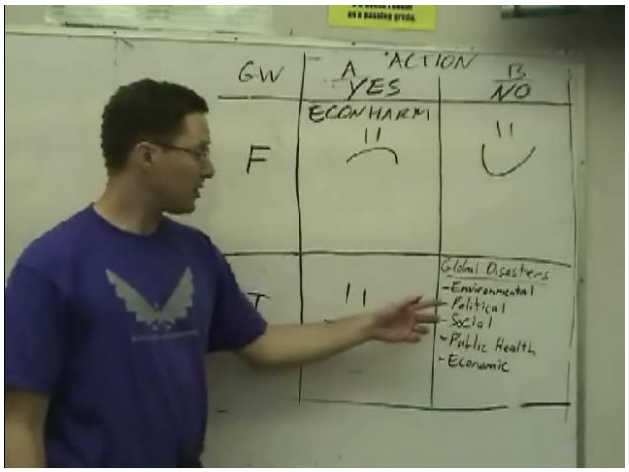


Figure 39: Yes Gesture, Yes Head (YGYH) Video

8.3.2.2 The No Gesture No Head Condition (NGNH)

In order to produce a similar condition without visible gesture and facial features, we use a video with only the diagram visible. The diagram is aligned with the video in the YGYH condition, so that the same AOIs can be used for all conditions. We use a diagram written on a digital whiteboard (Smartboard) rather than on a traditional whiteboard. In order to simulate the act of writing, we create movie clips with and without the writing and cut from one video to the next as the writing occurs in the original video. An image from the NGNH condition can be scene in Figure 40.

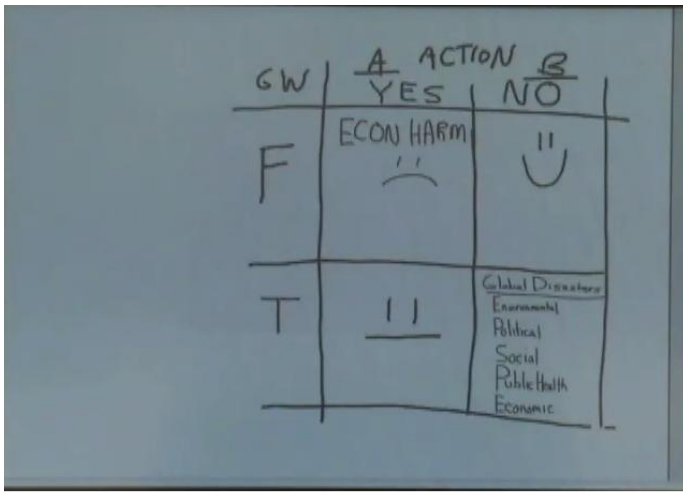


Figure 40: No Gesture, No Head (NGNH) Video

8.3.2.3 The No Gesture Yes Head Condition (NGYH)

To provide visual information about facial expression without communicating information about body language and gesture, we created a video that allowed the

presenter's head to appear overlaid on the underlying video as the presenter moved around in front of the whiteboard. By creating a "video mask" that leaves very little space around the presenter's head, it is possible to create a "disembodied" head moving around the video as the presenter interacts with the whiteboard. This technique has the benefits of the presenter's head appearing in exactly the same location as where the presenter's head appears in the YGYH video, allowing us to use the same AOIs for analysis across all conditions. Although we considered several other techniques for providing facial expression cues only (see Appendix 15.7), the floating head technique allows the most direct comparison to the other conditions and controls for the most extraneous variables (different head positions, potentially confounding duplicate artifact information being presented). An image from this video can be scene in Figure 41.

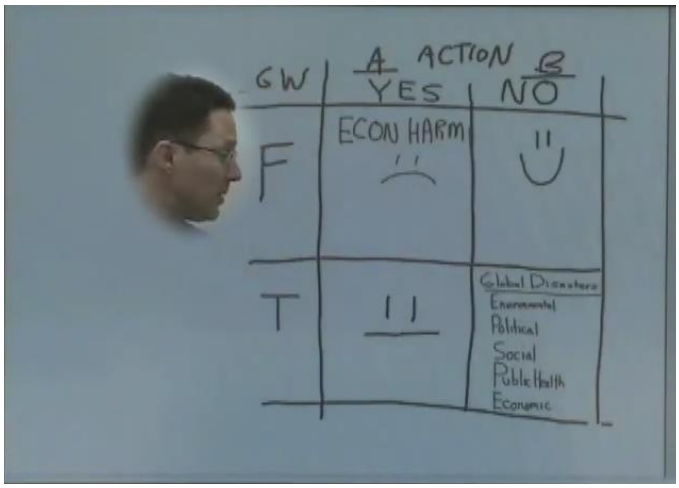


Figure 41: No Gesture, Yes Head (NGYH) Video

8.3.2.4 The Yes Gesture No Head Condition (YGNH)

To provide a gestural communication channel without adding a facial channel, we considered three possible mechanisms. First, the forearm of the actual presenter, as an overlay over the actual video, was considered. This uses a similar technique to that used for the floating head video in the NGYH condition. Second, we considered using a pointer icon with the length of the pointer approximately the length of the presenter's forearm. The goal was to simulate the location and orientation of the presenter's arm. Thirdly, we used a hand shaped icon that condition uses the NGNH video as its basis, with the overlay of the pointing icon tracking the location of the presenter's hand. The editing was again performed by a film student, resulting in high quality video overlays in

each case. In particular, a significant amount of time was devoted to creating realistic pointer trajectories that followed the presenter's hand.

The video overlay of the presenter's forearm was very difficult to do realistically (we wanted to avoid visuals that would distract from artifact interaction) and the size of the pointer looked unnatural when observed on its own. A pointer of similar size to the presenter's forearm also looked extremely artificial. We therefore chose to use the hand shaped pointer for the YGNH condition. An image from this video can be seen in Figure 42.

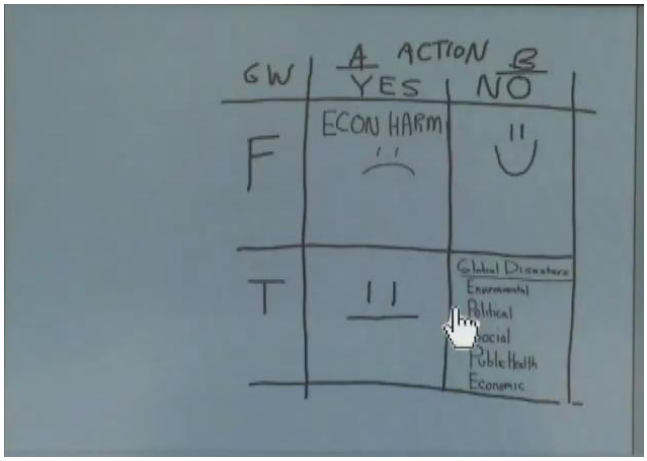


Figure 42: Yes Gesture, No Head (YGNH) Video with hand-shaped pointer

8.4 Participants

This dissertation is focused on studying how scientific researchers collaborate. We therefore draw the participants for our study from senior undergraduate students (4th year), graduate students, research associates, post-doctoral researchers, and faculty members. In addition, we specified that they should be active members of a research project. We accepted only subjects that fit the above criteria into the study. We recruited researchers from the University of Victoria's Faculty of Engineering, Faculty of Science, and Faculty of Social Sciences. We sent email recruitment letters (see Appendix 15.11 for the full text of the email) to departmental mailing lists and/or departmental administrators (who were asked to forward the email to their department). Subjects were offered a gift card for two movies at a movie theatre as remuneration for their participation.

Fifty subjects, 35 males and 15 females, took part in the study. The average age of the participants was 33 years, with a minimum age of 22 and a maximum age of 71. Thirty-four of the subjects were graduate students, ten were faculty, three were undergraduates, two were research associates, and one a post-doctoral researcher. The subjects spanned nine departments, twenty-three from computer science, nine from geography, five from biochemistry/microbiology, four from biology, three from electrical and computer engineering, two from electrical engineering, and one from each of mechanical engineering, math, and physics/astronomy. Thirteen subjects had to be rejected due to problems encountered with the accuracy of the eye tracking system (see Section 8.5.1 for details), leaving 37 subjects across the four conditions (10, 9, 10, and 8 subjects in the YGYH, NGNH, YGNH, and NGYH conditions respectively).

Subjects were blocked into groups of four, one for each condition. Conditions were assigned to subjects within a group randomly. We denoted a specific subject by specifying the group, the subject within the group, and the condition to which the subject was exposed. For example, the subject identifier prefix for the first group of four subjects appears as follows: 1-1, 1-2, 1-3, and 1-4. The random assignment of condition to subject can result in conditions being assigned to subjects in any order. For example, group one in the actual study consisted of the following assignment of conditions to subjects: 1-1-YGNH, 1-2-YGYH, 1-3-NGNH, and 1-4-NGYH. The randomization process resulted in the second group of subjects being assigned conditions in the following order: 2-1-NGNH, 2-2-YGYH, 2-3-NGYH, and 2-4-YGNH.

8.5 Study Apparatus

In this study, we are simulating a distributed presentation, with one of the remote observers (our subject) viewing the presentation from the desktop. The participants are provided with both audio and video from the presentation. The video is displayed on a 21 inch LCD monitor (LCD resolution of 1024x768) while the audio is played on standard desktop speakers.

A Tobii 2150 eye tracker is used to track subject's gaze. The eye tracker is carefully integrated into the 21 inch LCD, and is therefore not obtrusive to subjects. Head movement is relatively flexible with this system, with allowed movement within a region of 26 x 20 x 32 cm at a distance of 73 cm from the eye tracker. This allows for relatively

natural movement of the subjects while watching the presentation. The system uses near infra-red light emitting diodes to generate reflections in the subject's eyes and a large field of view camera to capture images and extract gaze. The user is not required to wear any special devices or hardware, and the system can track eye gaze for participants who wear glasses. The tracker is accurate to <0.7 degrees and has a spatial resolution of 0.35 degrees. It tracks at a rate of 50 Hz (a reading every 20 ms) with a latency of 35 ms.

We utilize the Tobii eye tracker to provide data about gaze fixations during the study. The Tobii analysis software (ClearView 2.7.1) allows an experimenter to change the gaze analysis parameters used in a given study. These parameters include filter settings (for data validity and how the two eyes are used to determine a fixation point), fixation radius (the number of pixels in which a fixation must remain to be a fixation), and the minimum fixation duration (the minimum time a fixation is allowed to be). This study utilizes gaze parameters similar to those recommended by Tobii for tracking reading on the screen. These parameters are the standard validity filter, averaging left and right eye to determine a fixation point, a 30 pixel fixation radius, and a 40 ms minimum fixation duration. We use these settings because the study utilizes text in many of the diagrams, and therefore reading is a common task in many scenes. For other tasks (looking at images, looking at mixed content), Tobii recommends a larger fixation radius and longer minimum duration. We need to maintain the short fixation duration so that a reading action where the subject is scanning across the screen can still be captured. Since we want our fixations to be accurate, we also want to maintain a relatively small fixation radius.

Before using the system for a specific subject, it is necessary to calibrate the tracking system. The calibration process is relatively quick, taking approximately 30 seconds, and is a robust process for most subjects. The calibration process consists of five circles appearing on the screen one after the other, with a subject asked to look at each circle as it shrinks to a single pixel. Since the eye tracker knows where the circles were drawn on the screen, it can confirm that the subject is indeed looking at the location on the screen. If the calibration is not accurate, the subject is asked to calibrate again.

Figure 43: Gesture Study Apparatus

An image of the eye tracking apparatus, as it was used in this study, is shown in Figure 43. This image shows an example subject (the author) watching the video. The tracker control screen is on the right of the image, at a ninety degree angle to the subjects screen and therefore not visible to the subject. A close up of the control screen is given in Figure 44. The control station allows the experimenter to monitor the tracking as it proceeds. Near the top of Figure 44, there is a small black square with two white circles, with a green bar at the bottom. This indicates that the tracker is tracking both of the subject's eyes (each circle represents an eye being tracked). In the middle of the screen, there is a blue circle and a blue line overlaid on top of the video, indicating the current gaze point of the subject as well as a trajectory indicating where the subject's gaze was in the recent past. The experimenter sat in the room with subjects as the experiment proceeded. Subjects were also video taped during the experiment. The video recording of the study recorded a similar view to that shown in Figure 43.

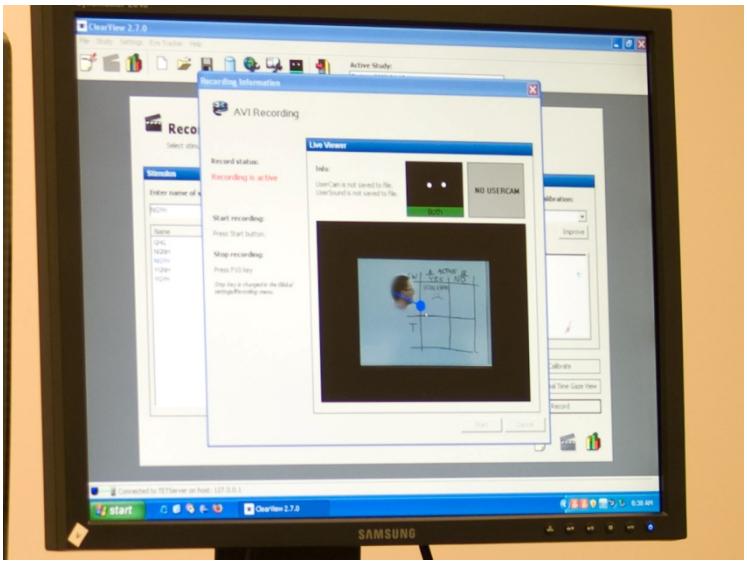


Figure 44: Gesture study control station close up

8.5.1 Tracking limitations

Although the Tobii tracking system is state of the art, we experienced several problems. These problems include the system not being able to calibrate for some subjects (due to strong prescription glasses, stigmatisms, or having a "lazy" eye) and the system introducing an offset into the tracking data part way through a study. The first case is simple to handle. Since the tracker cannot calibrate for a participant, that participant is excluded from the study at the very beginning of the study session. The second case is much more problematic. The problem arises when the calibration works, the system reports that it is tracking normally, but part way through a tracking session errors in the tracking data are clearly noticeable. The error is often large enough and consistent enough to be noticeable by the experimenter using the naked eye (e.g. rather than looking at the face, participants consistently look at a location above the head). As a result of this problem, we developed a post study calibration phase (Act 8 in the video) to detect such problems. Unfortunately, it was necessary to reject thirteen subjects as a result of this problem. More details on this issue can be found in Appendix 15.1.2.

8.5.2 Applying the CoGScience Framework to the Technology Domain

We use the CoGScience Framework to describe the technology characteristics of our study. From a study perspective, these characteristics are all control variables, as they are

kept constant across all of our conditions. Note that any one of the variables below could conceivably have been independent variables that we manipulated as part of the study (and indeed could be by future studies).

There are two **sensory streams** communicated to the study participants, an **aural stream** and a **visual stream**. The **aural stream** is a **high fidelity** stream and is played back on standard desktop computer speakers. Since the presentation is not distributed (the distribution is only simulated) there is no audio **feedback** or **delay**.

The visual stream is relative high **fidelity**, recorded at a **frame rate** of 30 frames per second and with moderate **clarity** (MPEG 4 codec). The **field of view** covered by the video varied depending on the scene, from fairly tight framing of the presenter filling a significant portion of the screen to a more wide angle framing that included the presenter and the whiteboard used in the video. Since the presentation is not truly distributed, **delay** and **quality** loss due to network loss and latency do not exist. The technology characteristics are discussed in more detail in Appendix 15.6.3.

8.6 Measurement and Observation

In this section, we outline the measures used to assess the effect of the treatments applied in this study. We use the CoGScience Framework to outline the measures at a high level. We then discuss the measures in detail. Finally, we end this section with a discussion of the instruments used for evaluation.

8.6.1 Using the CoGScience Framework to Determine Measures

The CoGScience Framework decomposes measures and outcomes into three distinct categories: **measures of process**, **measures of task**, **measures of group**, and **measures of cognition**. This study has two high-level goals: to determine if subjects attend to (look at) artifacts when watching a presentation and to determine if using gesture increases the understanding about the presentation topic. These two goals map directly onto the CoGScience categories of **measures of process** and **measures of task**. That is, the measures we utilize to explore how artifacts are observed are measures of the **communication process** itself. The measures we utilize to explore the impact gesture has on understanding the presentation topic are measures of the **task outcome** (to provide understanding). Since we do not have any true group interaction, we do not consider any

measures of the group process. Since our measures of the communication process (as described below) are also measures of attention, we also utilize **measures of cognition**.

8.6.1.1 Measures of Process

One of the goals of this study is to increase our understanding of what parts of a visual presentation researchers attend to during a research presentation. The main theoretical construct with which we are concerned is subject **attention**. That is, what are the subjects of the study looking at while they watch the presentation? Our hypotheses state that our interventions, whether gesture and facial expression are visible or not, will change what the subjects look at during the presentation. In particular, we hypothesize that artifacts will be attended to more often when gestural interaction is used to draw attention to those artifacts.

The dependent variables measured in this study revolve around **gaze fixations**. A gaze fixation occurs when a subject's gaze dwells on a certain area of the screen for more than a fixed amount of time (the gaze fixation threshold). Gaze fixations can be quite short (slightly higher than the gaze fixation threshold) or quite long (up to several seconds). A gaze fixation consists of a number of data elements, including the time the fixation started, the duration of the fixation, the screen coordinates where the fixation occurred, the scene in which the fixation occurred, the AOI in which the fixation occurred, and the type of AOI in which the fixation occurred. In a particular scene, many fixations often occur within the same AOI. We define the total fixation time within an AOI to be the sum of the durations of the fixations that occur within an AOI in a given scene (recall that each individual AOI only occurs in a single scene). This provides a measure of the importance of a specific AOI within a scene, and allows comparisons to be made against other AOIs in that scene as well as the time the subject's gaze is not fixated in any AOI.

Some scenes last for minutes, while others last only for seconds. AOIs are active as long as the scene in which they reside is on the screen. Thus, the percentage of time that a subject attends to an AOI provides a useful measure of the importance of an AOI relative to the scene duration. For example, the importance of a total fixation time of 400 ms within a single AOI is very different if that AOI resides in a scene with a duration of 500 ms (80% of the scene time) compared to a scene with a duration of 5000 ms (8% of the scene time). In addition, the ratio of fixation times between two different AOIs within a

single scene is also a useful measure. That is, a 2:1 ratio of fixation time between ExplicitPointArtifact and FacialFeature AOIs tells us that the ExplicitPointArtifact AOI is attended to more than the FacialFeature AOI in that scene.

In addition to these fundamental measures, we also consider a range of composite fixation based measures. These include:

- The total time and percentage of time spent within all AOIs in a given scene. This provides us with a measure of how well our AOIs capture the key elements of a scene.
- The total time and percentage of time spent within a given combination of AOI types. For example, we might be interested in the total and percentage of time spent in ExplicitPointArtifact, ImplicitPointArtifact, and ArtifactManip AOIs as AOIs that involve artifact interaction.
- We also consider several of these measures grouped across multiple scenes, across acts, across types of acts, and over the entire video. These groupings are defined as the sum of the AOI times (or the overall percentages) for a specific type of AOI across all scenes that occur in the grouping of interest. For example, we consider how a specific AOI type, such as ExplicitPointArtifact, is attended to across all of the scenes in each of our whiteboard acts, looking for effects from our treatment in a given act. We also consider this same gesture type across all of the whiteboard acts, looking for effects from our treatment across all whiteboard scenes. It is also possible to consider these measures across a subset of scenes within an act. Finally, for some types of AOI (e.g. ExplicitPointArtifact, ImplicitPointArtifact, ArtifactManip, and FacialFeature) we consider total time and percentage of time measures across the entire video.

8.6.1.2 Measures of Task

The goal of a research presentation (**task outcome**) is to communicate information to an audience, and in the case of the presentation used here, to convince the audience of the validity of the presenter's argument and that action should be taken against global warming. One of the goals of this study is to determine whether or not our interventions (the visibility of gesture and facial expression) have an impact on communicating

understanding. Thus, the dependent variables we are interested in measuring are primarily about understanding. In particular, we have three dependent variables:

1. Understanding about the artifact: The artifacts in the whiteboard acts are used extensively to support the presenter's argument. We want to measure the impact of our interventions on the subject's understanding of the artifact's **structure** (the different components of the artifact) and what structure components represent in the context of the presentation (e.g. what do the rows in the table represent in the Pascal's Wager diagram).

2. Understanding about the information presented using an artifact: The artifacts in the whiteboard acts present items of information that are critical to the presenter's argument. We want to measure the impact of our interventions on the subject's understanding of the **information** that is presented in the diagram (e.g. what information is presented in the bottom right quadrant of the Pascal's Wager diagram).

3. Understanding about the argument made using the diagram: Ultimately, the presenter's task is to communicate the point of the argument, using the whiteboard artifacts, as effectively as possible. We want to measure the impact of our intervention on the subject's understanding of the **argument** being made by the presenter (e.g. does use of artifacts help the subject understand the argument made by the presenter).

It should be noted that measuring understanding is quite difficult. We use questionnaires that measure the study participant's understanding of specific parts of the presentation that are important to the presenter's argument. It is the scores on these questionnaires that we use as a measure of understanding. The actual questionnaires are discussed in more detail in Section 8.7.2 and Appendix 15.13.

8.7 Data Collection

We use two primary mechanisms for collecting data about our study participants, eye tracking data and questionnaire data. The researcher was a passive observer during all sessions, and recorded relevant notes about each subject's activities during the session. Both the eye tracking software and the subject were monitored throughout the session

and any interesting or unusual events were noted. In particular, the eye-tracking software was monitored closely to ensure that the tracking was working throughout the session. An observer notes sheet was used for each subject and notes were made on this sheet (see Appendix 15.12). All subjects were videotaped, with the camera unobtrusively out of view of the subject (behind and to the right) while participating. The Tobii eye tracker was used to gather the process related eye tracking data, while pre-study, mid-study, and post-study questionnaires were used to gather the task related data.

8.7.1 Eye tracking data

8.7.1.1 Gaze fixations

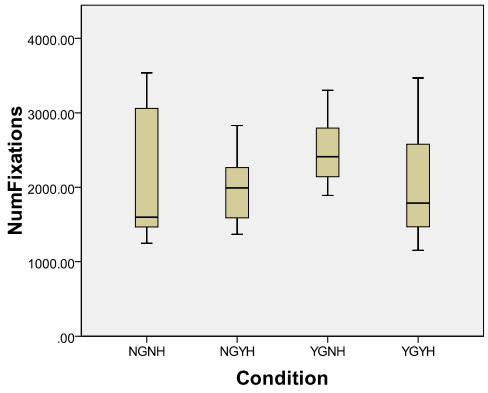


Figure 45: Number of fixations across the four conditions.

Before exploring gaze fixation in regards to specific AOI types, it is important to discuss the nature of the fixations that were recorded across the subjects in this study. The number of gaze fixations and the duration of those fixations are highly dependent on the subject. Some subjects have long, steady fixations while other subjects' eyes move rapidly and therefore have many short fixations. The number of fixations is highly variable, as can be seen in Figure 45, with a significant difference in the homogeneity of the variances but no statistical difference across the conditions. The average number of participant fixations across the entire 10-minute presentation is 2149, with a maximum of 3536 fixations (an average of 5.89 fixations per second) to a minimum of 1153 fixations (an average of 1.92 fixations per second).

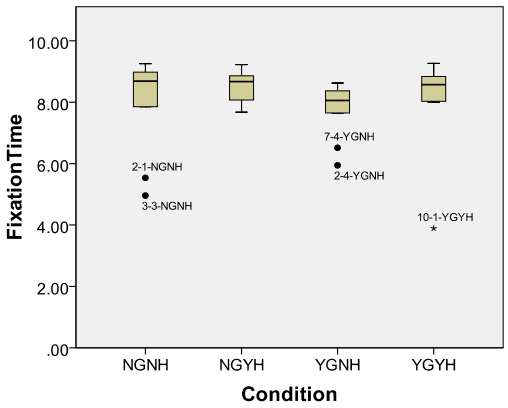


Figure 46: Total fixation times across conditions (in seconds)

Subjects spent an average of 8.1 minutes in a fixation state (fixated on something on the screen) over the 10-minute duration of the presentation, with a minimum total fixation time of 3.89 minutes and a maximum total fixation time of 9.27 minutes. These statistics are shown in Figure 46. Unlike the number of fixations, the total fixation time is not highly variable, with a relatively small number of outliers accounting for the extremes (3.89 to 9.27 minutes). The bulk of the participants have total gaze fixation times between 75% and 95% with 81% of the overall presentation (across all subjects in all scenes) spent in a fixation state.

Figure 47: Scene 1-1, Subject 10-1-YGYH with many short fixations

Given that there is a large variability in the number of fixations, it is important to consider what happens between fixations. We consider two extreme subjects. Subject 10-1-YGYH has a total time in a fixation state of 3.89 minutes (the most extreme outlier in Figure 46), with a large number of short fixations (3471 fixations with an average fixation time of 67 ms). The eye tracker detects short, jerky eye movements for this subject (Figure 47). Contrast this to subject 9-1-YGYH (Figure 48), with a total fixation

time of 9.27 minutes (the largest total fixation time) and a relatively small number of long fixations (1156 fixations with an average fixation time of 482 ms).

Figure 48: Scene 1-1, Subject 9-1-YGYH, a dialogue scene with AOIs and fixations

Since two adjacent fixations have a short gap between them, short jerky eye movements will typically result in shorter time periods in a fixation state (unless the time between fixations is very short). For subject 10-1-YGYH the gaps between fixations account for 6.1 minutes of the total presentation time of 10 minutes (3471 fixations at an average of 105 ms between fixations). Comparatively, subject 9-1-YGYH spends only 0.72 minutes of the ten minutes between fixations (1153 fixations at an average of 38 ms between fixations). It is important to point out that 10-1-YGYH's eye fixations are effective at capturing what subject 10-1-YGYH is attending to. That is, if you compare Figure 47 and Figure 48 it is clear that both subjects are primarily attending to the FacialFeature AOI and therefore both fixation "styles" capture the general nature of the participant's attention. This is important to note, as the overall fixation percentage time can be a misleading measure of fixation time if interpreted incorrectly. That is, it is not necessarily correct to say that Subject 10-1-YGYH pays less attention to FacialFeature AOIs than Subject 9-1-YGYH.

Although it is important that we are aware of this factor, it does not dramatically affect the results presented in our analysis. This is true for several reasons. First, fixation time has a relatively low variability, with most subjects spending 75% to 95% of the presentation in gaze fixation states (see Figure 46). Subject 10-1-YGYH is an outlier, and there are a small number of such outliers. Second, for the outliers that do have a significant amount of time in non-fixation states, fixations are typically grouped together

spatially with short times between fixations (on average 100ms). Although we cannot conclude that when in a non-fixation state the subject's eye gaze remains fixated in the region near the previous and following fixations, there is little to no evidence that gaze is fixated on other areas of the screen. That is, fixations are typically tightly grouped as with subject 10-1-YGYH. Third, our study design assists in dealing with this outlier problem, with the random assignment of subjects to our conditions theoretically balancing this problem across our conditions. Indeed, the analysis above (see Figure 46) does indicate that outliers are distributed across our conditions. And finally, we also take care to ensure that our interpretation of the impacts of gesture and facial expression visibility do not simply rely on the existence and duration of the fixations that are present in an AOI, but in addition comparatively analyze how fixations are distributed across multiple AOIs within a scene.

8.7.1.2 AOIs and fixations

Our AOIs are targeted at capturing gaze fixations on specific parts of the screen, typically facial expressions or artifacts. Each scene typically has a small number of AOIs, with the number of AOIs ranging from one to three per scene. In addition, the AOIs typically cover a small portion of the screen. See the Figures throughout Chapter 8, Chapter 9, and Chapter 10 for images of example scenes, including their AOIs. Only in rare cases where the presenter is making large sweeping gestures that cover a large part of the screen would the AOIs cover more than 15% or 20% of the screen. Even with such a small percentage of the screen covered by AOIs, a very large proportion of the fixation time is spent within AOI regions.

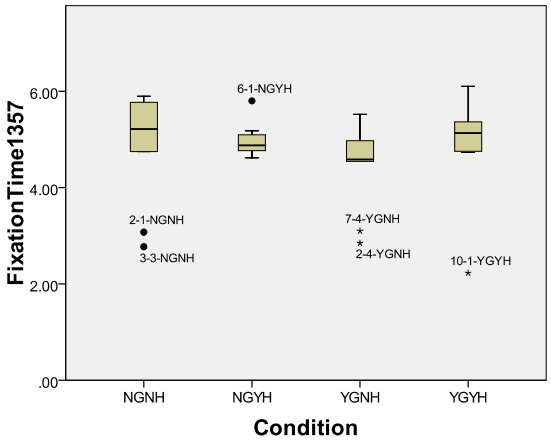


Figure 49: Total fixation time for Acts 1, 3, 5, and 7.

Across all of the Acts that do not include our experimental intervention (Acts 1, 3, 5, and 7), 70.84% of the total time is spent fixated within AOIs. Like total fixation time (see Section 8.7.1.1), the fixation time within AOIs in these acts has relatively low variability (see Figure 49), with most of the subject's fixation times falling between 4.5 to 6 minutes of the 6.8 minutes that these Acts span (66% to 88% respectively). There are also a small number of outliers as we saw with the total fixation time (primarily the same outliers in fact).

8.7.2 Questionnaires

Three questionnaires were used to gather data about the task-related aspects of the study. We briefly described these questionnaires below. The questionnaires are provided in their entirety in Appendix 15.13.1, 15.13.2, and 15.13.3.

8.7.2.1 Pre-study questionnaire

A pre-study questionnaire (see Appendix 15.13.1) was used to gather personal data about the subjects, such as age, gender, university department (e.g. Chemistry), position (undergraduate student, graduate student, research associate, post doctoral researcher, faculty, or staff), time in position, and research area. The questionnaire also asked about how often they attended presentations, how often they gave presentations, what tools and technologies (software and hardware) they used when giving presentations, and how often they used computers. The goal of this questionnaire was to gather demographic information, gather information about how familiar they were with attending and giving presentations, and gather information about their comfort level with computers.

8.7.2.2 Mid-study questionnaire

A mid-study questionnaire (as found in Appendix 15.13.2) was given to all subjects between the first video and the second video. This questionnaire presented the subjects with questions about the artifacts used in the presentation as well as the knowledge that the presenter was communicating to the audience. Recall that all subjects saw the same video, and that the purpose of this video and the questionnaire was to screen for strong individual differences that might help to explain outlier results in the main study. Because the only outliers detected in the study were outliers based on fixation duration as discussed above, this questionnaire was not used in our analysis.

8.7.2.3 Post-study questionnaire

A post-study questionnaire (as found in Appendix 15.13.3) was give to all subjects at the end of the second video. This questionnaire posed nine questions to the subjects about their understanding of the contents of the presentation as well as the structure and information contained in the artifacts used during the presentation (as described in Section 8.6.1.2). Below we present a high-level overview of questions asked in the post-study questionnaire. A more detailed discussion of the questionnaire can be found in Appendix 15.8.

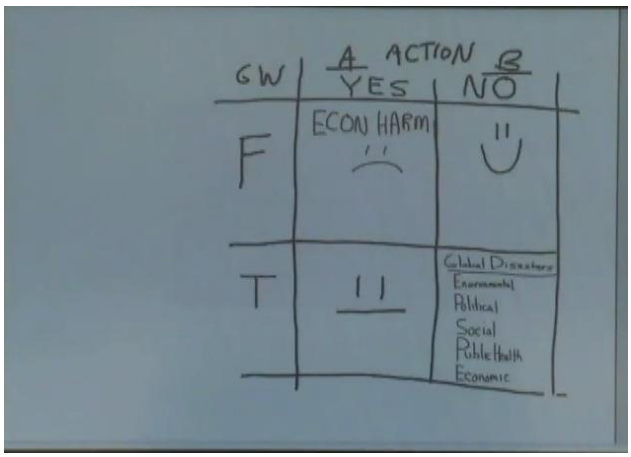


Figure 50: Pascal's Wager and Global Warming

Act 2 and Act 4 are two of the whiteboard scenes affected by our experimental conditions (gesture and facial feature visibility). Both acts contain extensive presenter interaction with artifacts on the whiteboard, in this case the Pascal's Wager diagram shown in Figure 50. It is on this diagram that we focus our questionnaire.

Question 1 poses an open-ended question that enables us to analyze qualitatively the subject's understanding of the use of the diagram and in particular, whether there is understanding that the diagram's primary role is to help make a decision when faced with uncertainty. It also allows us to determine whether a subject has been exposed to Pascal's Wager or a similar decision making tool in the past.

Question 2 and Question 3 deal with the *structural* components of Pascal's Wager diagram and the role they play in the presentation, and in particular what the roles of some of the diagram components play in the presentation. Question 2 asks *"What do the rows in the diagram represent?"* and Question 3 asks *"What do the columns in the diagram represent?"*

Question 4 and Question 5 deal with the *information* aspects of the diagram. Question 4 asks subjects to describe one of the key informational aspects about the presentation – "*According to the presenter, what are the potential risks of taking action against global warming? Which of these risks did the presenter list in the four quadrant diagram?"* We use this question because the answer (Economic Harm) is an artifact in the diagram (the top left quadrant in Figure 50), is created dynamically during the presentation (through writing), and is referred to repeatedly by gestures. Question 5 is similar to Question 4 except it measures understanding of a topic that is communicated by a different area of the diagram (the bottom right quadrant). Question 5 reads, "*According to the presenter, what are the potential risks of not taking action against global warming? Which of these risks did the presenter list in the four quadrant diagram?*"

Questions 6 and 7 measure a subject's understanding of the *argument* that the presenter is making. The logical argument posed by the presenter is that the evidence that humans are causing global warming is very strong, and therefore the bottom row is much more likely. In addition, the presenter attempts to demonstrate that the risks of not taking action are much worse than the risks of taking action. The questions asked are "*According to the presenter, which of the rows in the diagram is most likely to occur? What rationale does the presenter use to justify this position?*" and "*According to the presenter, which of the columns in the diagram has the most significant risk? What rationale does the presenter use to justify this position?*"

Question 8, like Question 1, is an open-ended question that allows us to qualitatively assess the overall understanding of the argument made by the presenter. "*According to the presenter, what is the key unknown in trying to understand what action we should take to deal with global warming?*" This question is designed to measure the subject's understanding of where the uncertainty lies in whether we should take action against global warming or not.

We use the answers to these questions to gauge the study participant's understanding. A more detailed discussion of the questionnaire, how the answers are scored, and the issues of using the questionnaire to determine understanding is given in Appendix 15.8. The process we used to ensure that the answers used in our analysis were unbiased is described in Appendix 15.9.3. The Questionnaire is presented in its entirety in Appendix 15.13.3.

8.8 Procedure

The study was held in the Usability Lab in the department of Computer Science at the University of Victoria. Subjects were asked to arrive at the lab at a specific time, with each subject allotted an hour to complete the study. Subjects typically took between 30 and 45 minutes to complete the study.

On arrival, the researcher explained the procedure that would be used in the study. Subjects were told that the study was exploring how researchers communicate about complex scientific topics and how computer technologies might be used to help researchers communicate more effectively. Subjects were told that they would be asked to watch two videos of a presentation about global warming and after each video they would be asked to answer a questionnaire about the content of the video. The first video was used to get all subjects familiar with the apparatus and the process. The second video contained our experimental interventions. Subjects were told that they would be video taped and that eye tracking would be used to track what they were looking at on the screen. Subjects were then presented with a consent form and asked if they had any questions. Once the consent form was signed, recording of the session commenced.

Subjects were first presented with the pre-study questionnaire and asked to fill it out. Once the questionnaire was completed, the researcher then explained the research

procedure in more detail. Subjects were told that before each video it would be necessary to calibrate the eye tracker and that the steps that would be taken would be:

- Calibrate the eye tracker;
- Watch the first video;
- Fill out a questionnaire about the content of the first video;
- Calibrate the eye tracker;
- Watch the second video; and
- Fill out a questionnaire about the content of the second video.

Since the Tobii eye tracker is limited in the range in which it can track effectively, it was necessary to explain to the subjects what types of movement were acceptable and what type of movement they should avoid. In particular, the problems encountered with tracker reliability (see Section 8.5) required that the subjects be asked to avoid certain types of movement (in particular slouching or leaning forward and backward). In order to relax the subjects with using the eye tracker, subjects were allowed to experiment with moving around in front of the eye tracker to see how the eye tracker responded to that movement. After this experimentation, subjects were asked to find a comfortable position to watch the first video. They were then asked to try to maintain that posture (and to continue looking at the screen) during calibration, during the time between the end of calibration and the start of the video (about 30 seconds), and during the first video's presentation. Subjects were reminded about the length of the video, the topic, and the fact that they would be given a questionnaire about the content of the video after the video was over.

All subjects went through the same calibration and video presentation during the first video. The researcher observed the subject and the eye tracker monitoring station, making notes about any interesting actions by the subject or events reported by the eye tracking system (e.g. subjects gazing off screen, subjects dozing off[7], possible tracking offset errors). Since the problems with offset errors reported by the tracking system, special

[7] Yes, one subject actually did doze off during the presentation. This demonstrates that the presentation is actually representative of a normal research talk. Isn't there always at least one person that falls asleep during a talk?

attention was taken in noting possible errors of this kind. After the first video ended, subjects were told to relax and were given the mid-study questionnaire.

After finishing the mid-study questionnaire, subjects were asked to once again find a comfortable position for watching the second video. Subjects were reminded that this video was different than the first video and that they would be given a second, different questionnaire at the end of the video. Subjects were also reminded that they should try and maintain their posture, that the second video was ten minutes long, that the topic of the video was global warming, and that the questions would be on the content of the video. The researcher observed both the subject and the eye tracking software, making notes as in the first video. At the end of the second video, subjects were told to relax and were provided with the post-study questionnaire. If subjects asked questions about either of the questionnaires, the researcher response was restricted to referring to the question on the questionnaire and asking the subject to answer the question to the best of their ability based on the content of the video.

8.9 Statistical Analysis Overview

In order to keep the statistical analysis in the next several chapters as concise as possible, we briefly outline our statistical approach. We follow the same procedure for all multi-factor analyses, although in some cases additional post-hoc analyses are performed. We summarize this information here as it is used throughout the analyses presented in the following chapters and the related appendices.

We test for both normality and homogeneity of variances to determine the types of tests to perform on our data. We use the Kolmogorov-Smirnov Z test [Nou04] to test for the normality of the distribution of our measures (H_o: the distribution is approximately normal, H_a: the distribution is non-normal). A Z statistic that is significant implies that we reject H_o and that our distribution is non-normal. A non-significant result implies that there is no evidence that distribution is non-normal. We use Levene's test [Lev60] to test for homogeneity of variance across the conditions (H_o: the variances across conditions are approximately the same, H_a: the variances across conditions are different). A Levene F statistic that is significant implies that the variances are different (we reject H_o), while a non-significant statistic implies that the variances are approximately equal.

If a set of measures are approximately normal and have homogeneous variances we use a General Liner Model based multi-factor ANOVA to perform the analysis [Dev82]. We use Tukey's HSD test to perform post-hoc pair wise tests [Dev82]. If the measure is approximately normal but the variances are not homogeneous we use a uni-variate ANOVA and utilize the Tamhane's T2 post-hoc analysis to compare pair-wise means [HT87]. The Tamhane T2 test (as opposed to Tukey's HSD) is not sensitive to non-homogeneous variances as long as the group sizes are approximately equal. If the measure we are analyzing is non-normal, we use the non-parametric Kruskal-Wallis test to perform a one-way ANOVA on the ranks of the samples [KW52]. In these cases we use the Mann-Whitney U test to perform pair-wise comparisons between the mean ranks of the groups [MW47]. All analyses are performed as two tailed analyses.

Our interpretation of our statistical results follows our pragmatic approach to obtaining knowledge (see Section 3.1.3). Although we use a baseline α level of 0.05 for the rejection of our null hypotheses, we do not consider our analyses as being black and white. Most of the results reported use an α level 0.05 but we recognize that results with a significance level between 0.1 and 0.05 may also be important to consider. Although we are careful to treat these results with the caution that they require, we do not dismiss them outright. When reporting results, we use phrases such as *"a significant difference"* to indicate significance levels in the range $p < 0.05$ and the phrase *"a moderately significant difference"* to indicate significance levels in the range $0.05 < p < 0.1$. All such statements are accompanied by their significance levels (p values).

The process described above is used throughout the following chapters to determine the types of analyses that are performed. Throughout the remainder of these chapters summary statistics only are provided. A detailed analysis of these statistics is provided in Appendix 15.14.

9 Understanding Gesture – Global Phenomena

We divide our analysis of the study described in Chapter 8 into three analysis chapters. In this chapter we consider gaze fixation phenomena in those acts in the video where there is no experimental intervention (all subjects see exactly the same presentation). This provides us with a broad understanding of what participants attend to in general before analyzing the impacts of our experimental intervention. Thus, we are measuring attention patterns for specific AOI types and in some cases comparing attention across different scene types. We call these the global gaze phenomena as they apply to all subjects within our study. Note that this analysis does not help us to address the hypotheses put forth in Section 8.2, but instead provides us with a basic analysis of how participants attend to facial feature and artifact AOIs in general. We leave the exploration of our experimental intervention (the impact of facial feature and gesture visibility), and therefore our hypotheses, to the analysis performed in Chapter 10.

9.1 Facial Expression

Facial expression, along with gesture, has long been recognized as one of the key non-verbal aspects of face-to-face communication (see Section 2.2.4). Although this study focuses on gestural interaction, it cannot do so at the expense of facial expression. Indeed, our experimental intervention controls for the visibility of both facial features and gesture. Before analyzing how facial features and gesture impact communication, we first consider how facial features are attended to in general. Acts 1, 3, 5, and 7 are all non-whiteboard scenes and therefore are not subject to our experimental intervention. In each scene, either the presenter or the Devil's Advocate is visible on screen, with a facial feature AOI enclosing the speaker's face. An example of such a scene, showing a FacialFeature AOI and two fixations, is shown in Figure 51. The fixation duration is denoted by the size of the circle, with the larger circle representing a fixation of 1794 ms and the smaller circle a fixation of 339 ms. The larger fixation is on the presenter's face (and inside the FacialFeature AOI) and the smaller fixation is on the bag of coal in the background of the scene.

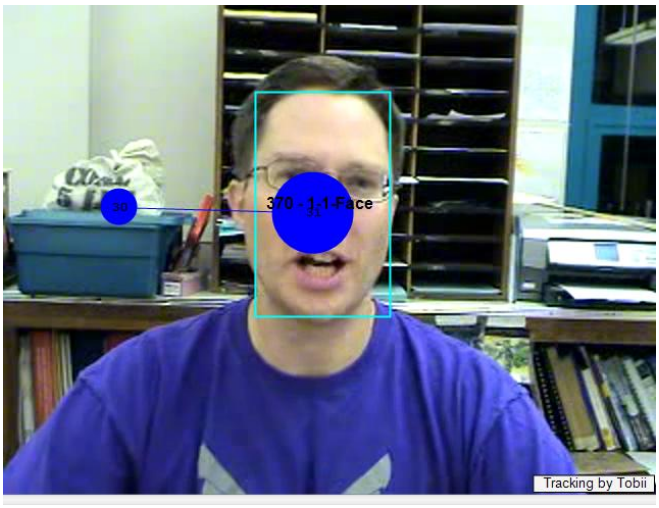


Figure 51: FacialFeature AOI with two fixations

Acts 1, 3, 5, and 7 have 87 scenes in total (see Table 5, p. 196), with each scene similar to that shown in Figure 51. Some of these scenes have just the presenter or the Devil's Advocate talking, while others have the presenter talking while using a physical artifact as a prop (see Figure 52). By considering the scenes in which no props are used (34 of the 87 scenes), we can analyze the attention paid to facial expression during simple dialogue.

Both the total scene duration and the total fixation time within FacialFeature AOIs are measured for all 34 scenes and all 37 subjects. In the 34 pure dialogue scenes, subjects spent 73.93% of the total scene duration time fixated on the FacialFeature AOIs. The minimum AOI fixation percentage for a single scene was 0% (some subjects, on some scenes, spent no time fixated on the facial features AOI) while the maximum facial feature fixation percentage for a single scene was 99% (some subjects, on some scenes, spent almost the entire scene fixated on the AOI). A pictorial representation of the measurements for a single subject in a single scene, as provided by the Tobii eye tracking software, is given in Figure 51 (Act 1, Scene 1 for subject 9-1-YGYH). The facial expression AOI fixation shown in this figure represents 74% of the total scene time for this subject.

9.2 Attending to Artifact Manipulation

We also measure gaze fixations in those scenes that involve physical artifacts, or props. There are three main sequences in the video in which props are used. One uses a paper version of Pascal's Wager as a challenge to the logic given by the presenter (2 scenes,

Figure 53), one uses a set of cans with paper signs that represent the potential impacts of global warming (21 scenes, Figure 52), and one uses a set of cardboard signs (3 scenes, not pictured). In each of these scenes, there are AOIs around the artifacts being manipulated (an ArtifactManip AOI) as well as an AOI around the speaker's facial features. In addition, when artifacts are manipulated quickly, we often create a post artifact manipulation AOI to capture fixations that lag behind the actual manipulation. Two example scenes of this type are shown in Figure 52 (Act 7, Scene 36, Subject 9-1-YGYH) and Figure 53 (Act 5, Scene 2, Subject 9-1-YGYH). There are two fixations in the first scene, a relatively long fixation in the artifact manipulation AOI (738 ms) and a relatively short fixation in the post artifact manipulation AOI (199 ms). There are many fixations in the second scene, with fixations within both the artifact manipulation AOI (paper figure) and the facial feature AOI.

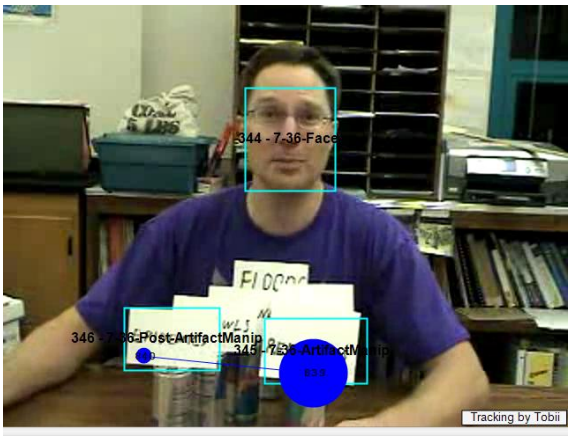


Figure 52: Dialogue scene with physical artifacts (cans) as props.

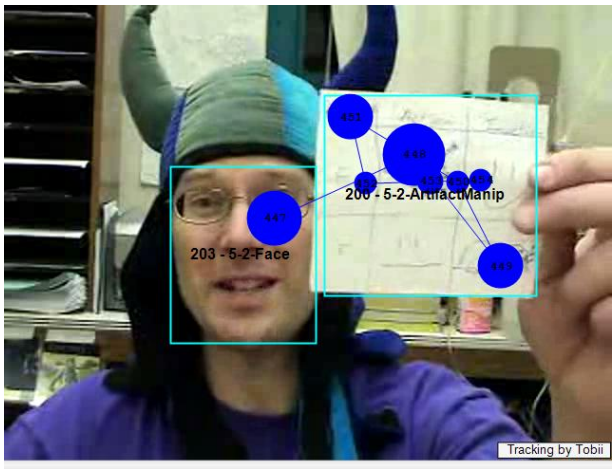


Figure 53: Dialog scene with physical artifact (paper diagram) as a prop

We measure both the total scene duration and the sum of the fixation durations within each AOI region of each artifact manipulation scene for all subjects. Subjects are fixated on the ArtifactManip AOIs 60%, 38% and 40% of the time for the paper Pascal's Wager, cardboard signs, and the cans sections of the presentation respectively.

We also measure the facial expression AOIs for theses scenes in the same way we did in Section 9.1. These measurements show that subjects are fixated on the FacialFeature AOIs for 13%, 26%, and 7% of the time for the paper Pascal's Wager, cardboard signs, and the cans sections of the presentation respectively. If we consider the FacialFeature AOIs across all three of these scene groupings combined, subjects are fixated on FacialFeature AOIs for 14% of the overall time.

If we consider the percentage of time each subject is fixated on facial features when there are no artifacts being manipulated, and compare that to the percentage of time subjects are fixated on facial features when artifacts are being manipulated, we can get an indication of whether or not artifact manipulation affects whether subjects attend to facial features. Both our simple analysis above and intuitive observations of the scenes depicted in Figure 51 and Figure 52 indicate that facial features are attended to less when artifacts are being manipulated. We explore this in more detail below.

In order to determine whether there is a difference between the attention paid to facial features when there are artifacts on the screen, we hypothesize that the average time spent in FacialFeature AOIs in the scenes in which artifacts are used will be smaller than those where no artifacts are used. For example, Subject 2-2-YGYH spends 22% of the total scene time in FacialFeature AOIs in those scenes where artifacts are used (29 scenes) and spends 74% of the total scene time in FacialFeature AOIs in those scenes where artifacts are not on the screen (35 scenes). We compare these percentages across all participants in our study. The percentage of time in FacialFeature AOIs for both artifact and non-artifact scenes is normally distributed (Kolmogorov-Smirnov $Z = 1.230$, $p = 0.097$ and $Z = 1.230$, $p = 0.933$ respectively)[8]. A paired samples (paired differences) t-test shows a statistically significant difference in the means, with a lower mean percentage of FacialFeature AOI time when artifacts are on the screen ($n = 37$, $t = 26.59$, $p = 2.94 \times 10^{-25}$). Thus, we can

[8] Recall that we are using a standard statistics process to test for normality, etc. Please refer to Section 8.9 for more details on the Kolmogorov-Smirnov Z statistic as a test for normality.

conclude that there is strong evidence that participants spend a significantly smaller amount of time attending to facial features when artifacts are being manipulated as part of the presentation.

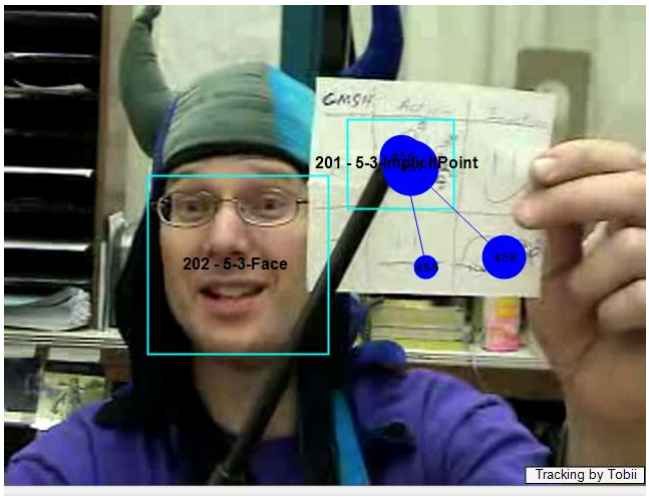


Figure 54: Dialogue scene with an implicit pointing gesture

9.3 Attending to Implicit Artifact Gesture

The analysis above considered a range of artifact interactions, the majority of which are artifact manipulations (moving the cans, holding up a sign). Although these types of artifact interactions are important, one of the key types of interaction in which we are interested is artifact pointing gestures. One key question that needs to be asked is do we observe a similar effect when considering such gestures. There are two scenes in the video where all subjects see the same implicit artifact pointing gesture. In these scenes, the Devil's Advocate is using an extreme example (the danger of giant mutant space hamsters) to show that Pascal's Wager can be used to make an argument for any case, no matter how ridiculous. In Act 5, Scene 4 (Figure 55 and Figure 56), the Devil's Advocate points at an image of the hamster eating a person, stating that it is better to build hamster traps *"…than even risk the possibility of being hamster chow"*. The artifact adds a visual aid to the argument and the implicit artifact pointing gesture is intended to draw attention to the part of the artifact about which the Devil's Advocate is speaking.

Note that in these scenes the AOI region has changed from encompassing the entire artifact when no artifact gesture is used (Figure 53) to encompassing the region around the part of the artifact to which the Devil's Advocate is pointing (Figure 54). The two

implicit artifact gesture scenes (Act 5, Scenes 3 and 4) for a single subject (Subject 9-1-YGYH) are shown in Figure 54 and Figure 55 respectively. A hot-spot analysis for Act 5, Scene 4 is shown in Figure 56. The hot-spot analysis represents the fixation counts across all of the YGYH subjects for this scene. Red areas of the image represent more than four fixations occurred in that region in this scene. Note the two prominent hot spots in the ImplicitPointArtifact and FacialFeature AOIs.

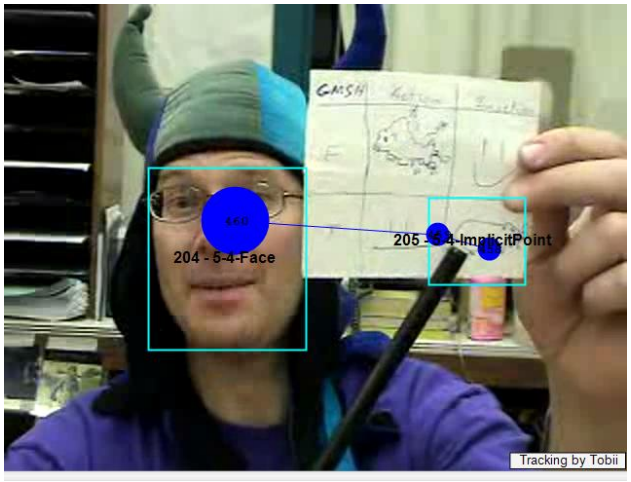


Figure 55: Dialogue scene with an implicit artifact gesture

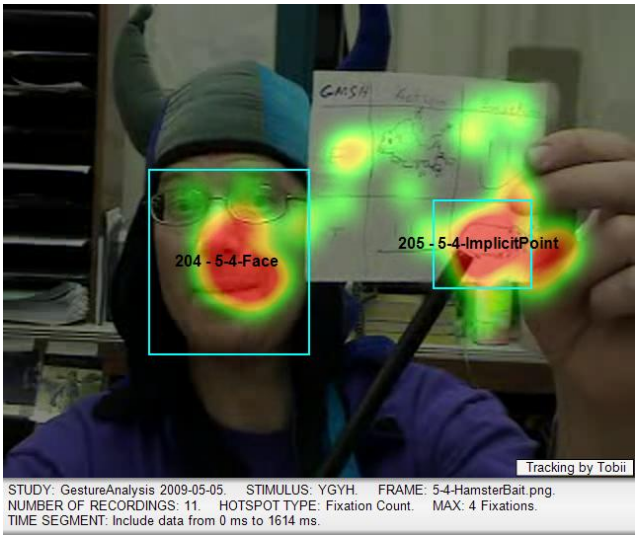


Figure 56: Hot spot analysis of fixation count for an implicit artifact gesture scene

Recall that because these scenes are not whiteboard scenes, we are not measuring the effect of our experimental intervention. Nor are we measuring the impact of gesture on attention to artifacts. Instead, we are measuring the difference in facial feature AOI

fixation time between scenes where implicit artifact gestures are used and when they are not.

We measure both the total scene duration and the sum of the fixation durations within each AOI region of the two implicit artifact gesture scenes for all subjects. On average, participants are fixated on the ImplicitPointArtifact AOIs 47% of the time. We measure the FacialFeature AOIs for theses scenes in the same way we did in Section 9.1. These measurements show that subjects are fixated on the FacialFeature AOIs for 13% of the time.

If we consider the percentage of time each subject is fixated on facial features when there are no artifacts being manipulated and compare that to the percentage of time subjects are fixated on facial features when implicit pointing gestures are being used, we can get an indication of whether or not implicit artifact gestures affect whether subjects attend to facial features. As when comparing facial feature attention in artifact manipulation scenes, both our simple analysis above and intuitive observations of the scenes depicted in Figure 51, Figure 54, Figure 55, and Figure 56 indicate that facial features are attended to less when implicit artifact gesture is being used.

The percentage of time spent attending to FacialFeature AOIs in the implicit artifact gesture scenes is not normally distributed (Kolmogorov-Smirnov $Z = 1.1483$, $p = 0.025$). Since the FacialFeature data is non-normal in the implicit artifact gesture scenes, we perform a non-parametric Wilcoxon Signed Rank Test (rather than a paired sample t-test) to test for a difference between the median ranks of the FacialFeature AOI time percentages. That is, we are considering whether the percentage of fixation time spent in facial expression AOIs is different in the sets of scenes that do and do not have implicit artifact gestures. The Wilcoxon Signed Rank Test indicates that the percentage of time spent fixating on FacialFeature AOIs is significantly smaller in implicit artifact gesture scenes than in non-artifact gesture scenes ($n = 37$, Wilcoxon $Z = -5.303$, $p = 1.14 \times 10^{-7}$). This provides evidence that there is significantly less fixation time spent on facial features in those scenes where implicit artifact pointing gestures are used.

9.4 Summary

The analysis presented in this chapter provides an important foundation for the broader discussion presented in Chapter 10. Recall that for the discussion in this section, we are

only considering those scenes where there is no experimental intervention (all subjects see the same video).

Result 1 Eye fixation is a good measure of subject attention.

The foundation of our measures is subject attention. Although attention is a broad topic (see Section 2.2.5.2 for a discussion), our focus is on attention that is activated by words and gestures. The cognitive psychology literature shows that when attention is drawn to an object through the spoken word (by asking someone to look at an object or manipulate it), gaze fixation rapidly focuses on the target object [TSE+95] as soon as the object is identified unambiguously. Given that our tasks require such attention (the speaker is asking audience members to attend to the artifact they point to), gaze fixation is a good measure of the effectiveness of gesture in drawing attention to artifacts. Our analysis of fixations during this study (Section 8.7.1.1) shows that 81% of the overall presentation time (across all participants in all scenes) is spent in an eye fixation state (participants are looking at something).

Result 2 Subject attention is captured effectively by the AOIs utilized in this study.

Our AOIs represent those areas of the screen that we anticipate study participants will attend to during the presentation. In fact, our AOIs are more restrictive, as artifact AOIs exist only if an artifact pointing gesture refers to that artifact. Given that the total fixation time within the AOIs used in this study is a significant percentage of the total presentation time (60%), our AOIs clearly capture "something" of interest within the presentation. Given that the AOIs capture only facial features or artifacts that have been either pointed at or manipulated, this percentage is quite high. With 60% of the presentation time fixated in our AOIs and 81% of the presentation time in fixation state, this leaves only 21% of the presentation time fixated on non-AOI areas of the screen. For example, any artifact on the screen that is not currently being pointed at by the presenter is not captured by our AOIs. Thus, if a participant fixates on any artifact on the screen other than one being indicated by a pointing gesture, that fixation would not be captured by one of our AOIs. Such fixations only accounts for 21% of the total presentation time. Since our AOIs cover a small percentage of the screen real estate while at the same time account for a significant percentage of the total scene time, this indicates that our AOIs are effective at capturing participant attention.

Result 3 ***Facial expression is attended to when a speaker is on the screen.***

Result 4 ***Artifacts are attended to when they are used as part of the presentation.***

Result 5 ***Artifacts are attended to when they are referred to as part of a gesture.***

Our analysis also suggests that FacialFeature AOIs are attended to a significant proportion of the time (74% of the total time in those scenes where artifacts are not utilized), that artifacts are attended to when they are used as part of the presentation (42% of the presentation time in those scenes where artifacts are used), and that artifacts are attended to when they are referred to by a pointing gesture (47% of the presentation time in those scenes where implicit artifact gestures are used). Note that it is important to be clear that these results do not explore the effect of gesture or facial feature visibility on attention, only that participants attend to FacialFeature, ArtifactManip, and ImplicitPointArtifact AOIs.

Result 6 ***Subjects do not attend to facial features as often when artifacts are a part of the presentation.***

Result 7 ***Subjects do not attend to facial features as often when implicit artifact gestures are used as part of the presentation.***

Finally, we show that there is a statistical difference in the amount of time spent attending to FacialFeature AOIs when artifacts are being manipulated on the screen. This holds true for both artifact manipulation as well as artifacts that are pointed at using implicit artifact gestures.

These results show that both eye fixation and the AOIs defined for this study appear to be a valid mechanism for acquiring a measure of subject attention for both gestural interaction and facial expression. In addition, these results indicate that there may be an interesting interaction between the visibility of facial expression and gesture. We explore this interaction in more detail in the following chapters.

10 Understanding Gesture: Experimental Intervention

Chapter 9 looks at measures that are consistent across all subjects, as it only considers those scenes in which there is no experimental intervention. In this chapter, we explore the impact of our experimental interventions on gaze fixation. In particular, we consider the visibility of facial features and gesture as the independent variables. We analyze the results of two types of measures in this chapter. First, we consider measures of collaboration process (Section 10.1). These are the measures that help us to determine whether researchers attend to artifacts in general as well as whether or not artifact gestures assist in drawing attention to those artifacts. These analyses provide us with results that will either support of refute our hypotheses that involve artifact attention:

- *Hypothesis 1: Researchers will attend to artifacts when they are used as part of a presentation.*
- *Hypothesis 2: Researchers will attend to artifacts more frequently when gesture is used to draw attention to an artifact.*
- *Hypothesis 4: Researchers will attend to facial expression when it is communicated as part of a presentation.*
- *Hypothesis 5: Researchers will attend to the artifacts used in a presentation less when facial expression is visible as part of the presentation.*

Ultimately, a scientific presentation is about communicating concepts and information clearly. The analysis of our task measures (Section 10.2) helps us determine whether gesture and facial feature visibility have an impact on the understanding of the research presentation. These analyses provide us with results that will either support or refute our hypotheses that involve understanding:

- *Hypothesis 3: Researchers will have a better understanding about artifacts, how they are used, and the information they contain when gesture is used to refer to those artifacts during a presentation.*
- *Hypothesis 6: Researchers will have a poorer understanding about artifacts, how they are used, and the information they contain when facial expression is visible as part of the presentation.*

We break down our detailed statistical analysis into two sections. Our low level analysis considers the impact of facial feature and gesture visibility on AOI attention for each gesture type (emphatic, explicit, implicit, and manipulation) across each AOI type. This is a relatively mechanical process, analyzing the impact of our intervention on attention to each of the AOI types for each artifact gesture type using the statistical techniques described in Section 8.9 (ANOVA and appropriate post-hoc tests). This analysis is presented in detail in Appendix 15.14 and the reader is referred to the Appendix if interested in these details.

In this section, we consider the results from the detailed analysis presented in Appendix 15.14, synthesizing the individual analyses for a specific gesture or AOI type into a comprehensive analysis across the different types of gesture studied. The primary goal of this chapter is to distil our analysis into a concise set of research results.

10.1 Measures of Process

In this chapter we consider four types of communication events, exploring their efficacy in drawing attention to artifacts as well as their impacts on the attention to other AOIs. These communication events are emphatic gesture events, implicit artifact events, explicit artifact events, and artifact manipulation events. Fundamental to this study is the use of AOIs that allow us to analyze the impacts of our experimental intervention (gesture and facial expression visibility) on our study participants' attention to artifacts and facial features. We consider these dimensions below by considering the impacts of our intervention on EmphaticGesture, Artifact, FacialFeature, and total AOI fixation times. For more information on the definition of these communication events, see Section 7.1.3 of our Ethnography. For more details on the definitions of the AOI types, see Section 8.3.1.3.

In the following sections, we divide up the presentation of our statistical analysis based on the impacts our experimental interventions have on attention paid to EmphaticGesture AOIs, artifact AOIs (ExplicitPointArtifact, ImplicitPointArtifact, and ArtifactManip AOIs), and FacialFeature AOIs. Note that this analysis is different than the analysis presented in Appendix 15.14.1. In Appendix 15.14.1, we consider the impacts of facial feature and gesture visibility for each gesture type (emphatic, implicit artifact, explicit artifact, and artifact manipulation) individually. That is, we look at the impact of facial

feature and gesture visibility on AOI attention for each gesture type across all AOI types but do so in isolation of the other artifact gesture types. This is a relatively mechanical process, considering each artifact gesture type in turn, and for each artifact gesture type we consider the impact of our intervention on attention in each of the AOI types.

In this section, we focus our analysis on the impact of facial feature and gesture visibility on the attention paid to specific AOI types across all types of artifact gestures. This analysis combines our individual analyses of the gesture types from Appendix 15.14.1 into a comprehensive analysis of the impacts of facial feature and gesture visibility on attention to emphatic gesture AOIs (Section 10.1.1), artifact AOIs (Section 10.1.3), and facial feature AOIs (Section 10.1.4). It is this analysis that provides us with results that either support or refute the artifact-centric hypotheses discussed above.

10.1.1 Impacts of facial feature and gesture visibility on EmphaticGesture AOIs

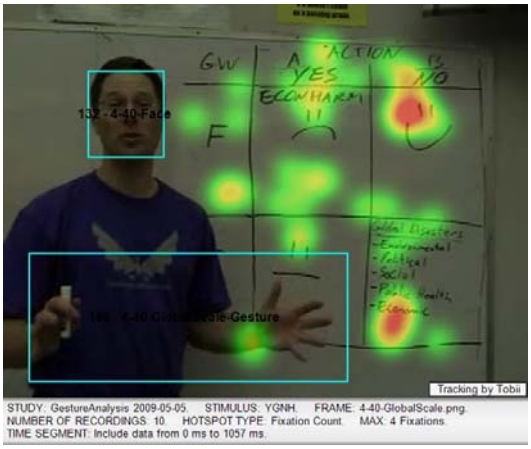


Figure 57: An Emphatic Gesture with hot spot analysis

Emphatic gestures are those gesture/utterance pairs that are used for emphasis but do not play a role in artifact interaction. An example of such a gesture, including a hot-spot analysis for the YGNH condition, is shown in Figure 57. There are nine scenes in Act 4 in which emphatic gestures occur. We measure fixation time within the EmphaticGesture AOI for each subject in each of the nine scenes that contain an emphatic gesture. Next, we aggregate the EmphaticGesture fixation times for each subject across the emphatic gesture scenes. We then analyze the total fixation times per subject across our experimental conditions (gesture and facial expression visibility).

Result 8 ***In emphatic gesture scenes the YGNH condition (gesture visible, facial feature not visible) has significantly higher EmphaticGesture AOI fixation times than the other conditions (YGYH, NGYH, and NGNH).***

Result 9 ***In emphatic gesture scenes our analysis provides no evidence that gesture visibility has an impact on fixation times in EmphaticGesture AOIs between the YGYH, NGYH, and NGNH conditions, on fixation times in FacialFeature AOIs, or on the total fixation time across all AOI types.***

The detailed analysis of EmphaticGesture (see Appendix 15.14.1.1) provides us with several interesting results. First, emphatic gesture is attended to significantly more when gesture is visible and facial features are not visible (the YGNH condition versus YGYH, NGNH, and NGYH, Tamhane post-hoc $p = 0.025$, $p = 0.025$, and $p = 0.023$ respectively). In the conditions where gesture is not visible (the NGYH and NGNH), there is no significant difference in the mean attention to emphatic gesture. This is not surprising, as the emphatic gesture is not visible in both conditions. Surprisingly, even with visible gesture in the YGYH condition (where facial features are also visible), there is no statistically significant difference between this condition and either the NGNH and NGYH conditions. Note that gesture visibility has no significant impact on either FacialFeature or Total AOI fixation times. This implies that in emphatic gesture scenes, our results provide no evidence that gesture visibility has an impact on AOI fixation except when gesture is visible and facial features are not.

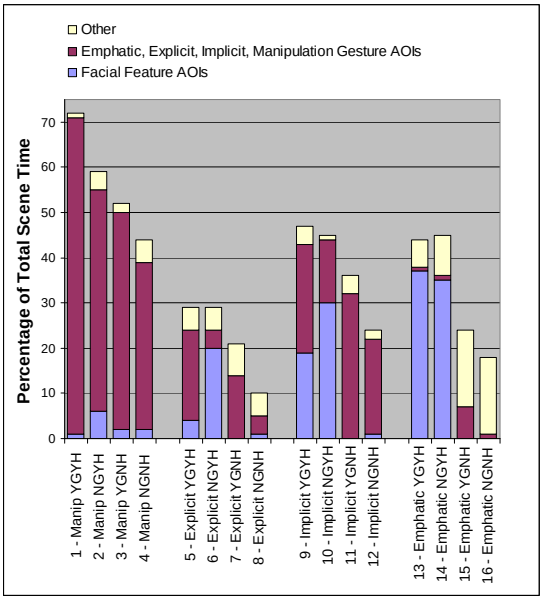


Figure 58: Percentage of total fixation time for all AOI types in gesture related scenes

Result 10 Emphatic gesture AOIs are not attended to at the same level as artifact manipulation, explicit artifact, and implicit artifact AOIs except in the YGNH condition.

Another key result from this analysis is that subjects do not attend to EmphaticGesture at the same level as they do to other gesture types. The percentage of fixation time in EmphaticGesture AOIs (Columns 13 – 16 in Figure 58) is almost non-existent (1% of the total time) except in the YGNH condition (7% - Column 15). Contrast this to the approximately 35% of FacialFeature AOI fixation time when facial features are visible in these same scenes (Column 13 - 16). Note that the total time spent in fixation for emphatic gesture scenes is comparable to other scene type. This analysis implies that participants do not attend to emphatic gesture to the same degree that they attend to other gesture types except in the YGNH condition.

10.1.2 Impacts of facial feature and gesture visibility across all AOI types

The other three types of gesture we consider are fundamentally different from emphatic gesture. Emphatic gestures are utilized to emphasize a verbal point being made by the presenter, while implicit artifact, explicit artifact, and artifact manipulation events all refer to and/or modify the artifacts that are being used. That is, their referent has a digital representation. If we accept the fact that a presenter pointing at an artifact implies that he/she is trying to draw the attention of the audience to that artifact (why else would he point to it, after all), what can our analysis tell us about the efficacy of these types of gestures in performing such a function? This analysis provides us with evidence that either supports or refutes two of our primary hypotheses: *Researchers will attend to artifacts more frequently when gesture is used to draw attention to an artifact.* and *Researchers will attend to the artifacts used in a presentation less when facial expression is visible as part of the presentation.*

		Two-Way ANOVA (n = 37)			One Way ANOVA (n = 37)
Scene Type	AOI Type	Gesture Visibility	Facial Visibility	Interaction	
Implicit Artifact	Artifact	$p = 2.29 \times 10^{-5}$	$p = 0.002$	$p = 0.847$	
	Facial				$p = 4.0 \times 10^{-15}$
	Total	$p = 0.026$	$p = 5.28 \times 10^{-6}$	$p = 0.073$	
Explicit Artifact	Artifact	$p = 8.12 \times 10^{-7}$	$p = 0.44$	$p = 0.695$	
	Facial				$p = 5.19 \times 10^{-10}$
	Total	$p = 0.06$	$p = 1.38 \times 10^{-5}$	$p = 0.044$	
Manipulation	Artifact	$p = 0.001$	$p = 0.001$	$p = 0.26$	
	Facial				$p = 2.36 \times 10^{-6}$
	Total	$p = 0.045$	$p = 0.001$	$p = 0.588$	

Table 6: Analysis of Variance Summary Statistics

The analysis presented in Appendix 15.14.1 considers the effects of gesture visibility and facial feature visibility on artifact AOIs (ImplicitPointArtifact, ExplicitPointArtifact, and ArtifactManip), facial feature AOIs, and total AOI fixation time. These results are summarized in Table 6. We perform this analysis for each of the three artifact focussed scene types within the video (implicit artifact gesture, explicit artifact gesture, and artifact manipulation scenes). That is, we first consider the effect of gesture and facial feature visibility on artifact fixation time, facial feature fixation time, and total fixation time in those scenes where implicit artifact gestures occur (the top three lines in Table 6). We then perform the same analysis in those scenes where explicit artifact gestures occur (the middle three lines in Table 6), followed by an analysis in the artifact manipulation scenes (the bottom three lines in Table 6).

We follow the statistical procedure outlined in Section 8.9, using a two-way multi-factor ANOVA for normally distributed measures with homogeneous variances and a one-way ANOVA for normally distributed measures with non-homogeneous variances. Our detailed ANOVA analyses are provided in Appendix 15.14.1.

Result 11 In almost all scene types and for almost all AOI types, there is a significant main effect of gesture visibility on AOI attention.

Result 12 In almost all scene types and for almost all AOI types, there is a significant main effect of facial feature visibility on AOI attention.

The above analyses indicate that there are main effects for both gesture and facial feature visibility on the mean fixation time in almost all AOI types in all scenes ($\alpha = 0.05$). Thus, our experimental interventions have a clear impact on fixation time. There

are only two cases where there are no significant results at the $\alpha = 0.05$ level. The significance level of gesture visibility on the total AOI fixation time in explicit artifact gesture scenes is moderately significant ($p = 0.06$) while the significance level of facial feature visibility on fixation time in artifact AOIs in explicit artifact gesture scenes is not significant ($p = 0.44$).

Although this analysis shows us that both facial feature and gesture visibility have a strong impact on fixation time in our AOIs, this is only a high-level analysis of the impact of our experimental interventions. It does not provide us with a detailed analysis of which conditions are affected and how. This analysis is provided through post-hoc analyses. As described in Section 8.9, we use the Tukey HSD pair-wise test for those measures that have homogeneous variances and a Tamhane T2 post-hoc pair-wise test for those measures that have non-homogeneous variances. Our detailed analysis of AOI fixation times for each gesture type (implicit artifact, explicit artifact, and artifact manipulation) and for each type of AOI fixation time (artifact, facial feature, and total AOI time) listed in Table 6 is given in Appendix 15.14.1. In this section, we synthesize our analyses of how the three gesture types impact attention for a specific AOI type.

We first consider the impacts of our experimental interventions on attention to artifact AOIs (Section 10.1.3) across all three artifact gesture types (implicit artifact, explicit artifact, and artifact manipulation). We then perform the same analysis on facial feature AOI fixation time (Section 10.1.4), and total AOI fixation time (Section 10.1.5). We also consider the effectiveness of artifact gestures in drawing attention to artifacts (Section 10.1.6). As we did in Chapter 9, we present key findings from these analyses listed as experimental results, highlighting interesting findings in bold. As with all of our statistical analyses for this study, a more detailed treatment of each statistical test can be found in Appendix 15.14.1.

10.1.3 Impacts on artifact AOIs across gesture types

10.1.3.1 Artifact AOIs and gesture type

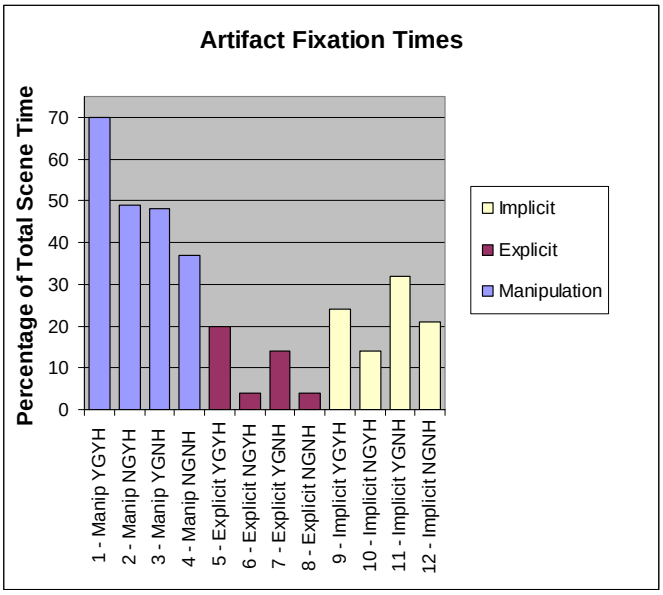


Figure 59: Percentage of total fixation time for artifact AOIs

Result 13 Artifact manipulation scenes (across all conditions) have the largest percentage of overall fixation time.

Initially, we consider the percentage of total fixation time for each type of the artifact interaction events. Figure 59 reveals several interesting results. First, the ArtifactManip AOIs clearly have the largest percentage of total fixation time, with the lowest ArtifactManip AOI percentage higher than the highest percentage of either the ImplicitPointArtifact or ExplicitPointArtifact AOIs. Although why this occurs is not clear, one conjecture is that it is the interactive or dynamic nature of the action that draws attention to the artifact. It is tempting to attribute this to the gesture itself, but even in the NG artifact manipulation conditions (where there is no visible gesture and the text/drawing just appears) the artifact fixation percentages are quite high. Thus, it appears that it is both the artifact manipulation gesture and the manipulation of the artifact (the text being written, a line appearing on the screen) that draws and maintains the attention of the subjects during such scenes.

Result 14 The non-visible gesture conditions of the explicit artifact gesture scenes have the lowest overall artifact fixation times.

At the other end of the spectrum, we see that the NG conditions of the explicit artifact scenes have a smaller total fixation percentage than any of the other conditions. Recall

that an explicit artifact gesture is deictic in nature, and that when subjects do not have visible gesture there is no way to determine the referent. That is, without the gesture, subjects have no idea what to look at. In some cases, the context of the speech might give enough of a clue to be helpful. For example, the phrase "what about this box down here" has enough context in the phrases "this box" (telling us the presenter is talking about a box) and "down here" (that the box is near the bottom of something) to direct the subject's gaze to the correct area in the diagram. However, most of the time this is not the case and this is shown clearly by the small percentage of total fixation times in the explicit artifact gesture conditions where gesture is not visible.

Result 15 The explicit artifact gesture scenes have the lowest overall artifact fixation times.

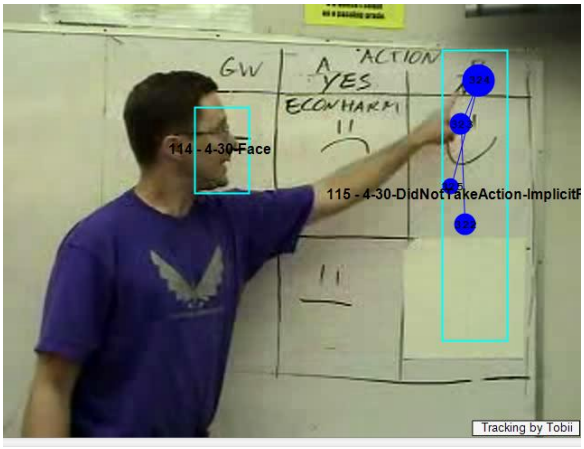


Figure 60: An ImplicitPointArtifact event

In addition, the explicit artifact gesture scenes also seem to have the lowest average fixation percentage of the three artifact gesture scene types. Although why this might be so is unclear, it is worth noting an important difference between how the presenter uses explicit and implicit artifact gestures in the video. In many cases, the presenter uses implicit artifact gestures to refer to fairly large AOI regions. An excellent example is given in Figure 60 where the presenter is referring to an entire column of the Pascal's Wager diagram. In general, the presenter uses explicit artifact gestures to refer to much smaller areas of the screen, such as a single quadrant of the diagram.

10.1.3.2 Artifact AOIs and gesture visibility

We next consider how gesture visibility impacts fixation time across the artifact gesture types (manipulation, explicit, and implicit artifact interactions). We perform pair-wise

comparisons between conditions that differ in their gesture visibility (we vary the G in the condition) but are the same in their facial expression visibility (we keep the H constant). For example, we consider the pair NGNH and YGNH as conditions where facial feature visibility does not change (constant NH). The percentage of total fixation times for the three artifact gesture AOI types are given in Figure 59 and the conditions we are comparing across are given in Table 7. Note that the percentages in Figure 59 are the same as the "Gesture" percentages in Figure 58 and that the "Gesture" category in Figure 59 represents the ArtifacManip, ExplicitPointArtifact, or ImplicitPointArtifact AOIs, depending on the type of scene listed on the X axis.

Artifact AOIs	NG	YG	Mean Difference	p-value	Test
ImplicitPointArtifact Scenes	NGNH	YGNH	5937	0.004	Tukey
	NGYH	YGYH	5487	0.012	Tukey
ExplicitPointArtifact Scenes	NGNH	YGNH	1561	0.002	Tukey
	NGYH	YGYH	1779	0.0005	Tukey
ArtifactManip Scenes	NGNH	YGNH	2031	0.342	Tukey
	NGYH	YGYH	4003	0.014	Tukey

Table 7: Pair-wise comparisons of Artifact AOIs (varying G, constant H).

Result 16 Gesture visibility has a significant impact on the attention paid to Artifact AOIs.

Examining these pairs across all gesture types provides a broad view of the impact of gesture on all types of artifact fixations. We use the Tukey HSD post-hoc test to do mean difference pair-wise comparisons, the results of which are given in Table 7. For five of the six comparisons (three types of gesture, each with two pairings per gesture type), there is a significant increase in the mean gesture based AOI fixation time (the mean difference is positive). These trends are also visible in Figure 59 as increases in the percentage of Gesture time from Columns 2 to 1, 6 to 5, 10 to 9, 8 to 7, and 12 to 11. They are also visible in the estimated mean plots for the implicit, explicit, and manipulation artifact gestures as given in Appendix 15.14.1 and in pictured in Figure 61. Only in the artifact manipulation NGNH/YGNH comparison (Columns 4 to 3 in Figure 59 and the bottom line in the rightmost graph in Figure 61) is the increase not statistically significant.

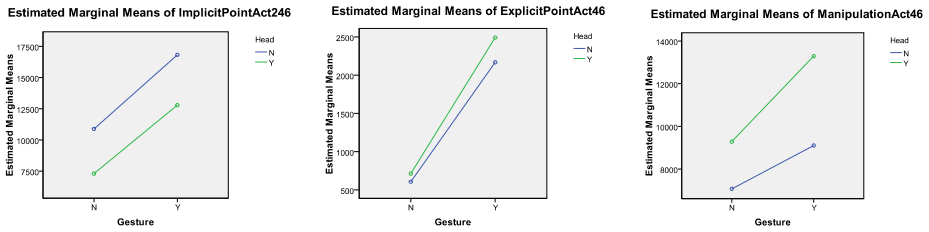


Figure 61: Estimated Marginal Means for Implicit, Explicit, and Manipulation Gestures

Although the reason why the artifact manipulation NGNH/NGYH pairing is not significant is unclear, there is one important factor to consider. Artifact manipulation gestures attract a significant amount of fixation time, even when there is neither gesture nor facial feature visibility (37% of the total scene time in the NGNH condition). The highest percentage of artifact fixation time in the explicit artifact and implicit artifact scenes is the implicit artifact gesture YGNH condition at 32%. Given that the artifact manipulation scenes already have such a high percentage of total scene time fixated in the artifact AOI (the lowest artifact manipulation percentage being higher than the highest percentage of the other artifact interaction types), artifact manipulation is clearly effective at drawing attention to artifacts without any gesture being present at all. Another possible explanation is that because manipulation draws significant attention to the artifacts without either gesture or facial expression visibility, our low fidelity pointer may not be a compelling enough visual cue to attract more attention.

Artifact AOIs	NH	YH	Mean Difference	p-value	Test
ImplicitPointArtifact Scenes	NGNH	NGYH	-3581	0.175	Tukey
	YGNH	YGYH	-4030	0.068	Tukey
ExplicitPointArtifact Scenes	NGNH	NGYH	106	0.994	Tukey
	YGNH	YGYH	324	0.821	Tukey
ArtifactManip Scenes	NGNH	NGYH	2215	0.316	Tukey
	YGNH	YGYH	4186	0.006	Tukey

Table 8: Pair-wise comparisons of Artifact AOIs (varying H, constant G).

10.1.3.3 Artifact AOIs and facial feature visibility

Result 17 Visible facial features result in a significant increase in attention on artifact AOIs in artifact manipulation scenes when gesture is visible (YGNH/YGYH).

When we consider comparisons between facial feature visibility (varying H) across constant gesture conditions (constant G), the results are not consistent across gesture types (Table 8). In the artifact manipulation YGNH/YGYH comparison, there is a significant increase in the mean artifact AOI fixation time ($p = 0.006$). Indeed, it is

tempting to assume that the condition that is closest to face-to-face (YGYH) would result in the highest mean fixation and percentage of scene time in the artifact based AOIs. Although this is true in the artifact manipulation and explicit artifact gesture scenes, it is not the case in the implicit artifact gesture scenes. In the implicit artifact gesture scenes, the YGNH condition has a significantly higher ($p < 0.05$) mean than NGNH and NGYH conditions (not shown in the table) and has a higher ($p = 0.068$) mean than the YGYH condition (See Table 8 and Figure 59). This is essentially the inverse result as that observed in the artifact manipulation scenes. Although the ANOVAs summarized in Section 10.1.2 (for details see Section 15.14.1.2) do not show a statistically significant interaction effect between gesture and facial expression visibility, these pair-wise results are nonetheless interesting. It is also worth noting that in the implicit artifact and explicit artifact scenes, the NGYH condition often has a lower or similar mean to the NGNH condition. This implies that having visible facial features rarely significantly increases the mean fixation time in artifact AOIs (with the exception of the artifact manipulation YGNH/YGYH case), in many cases there is no evidence that it has any significant impact at all, and suggests that there is evidence that they may in fact have an adverse impact on attracting attention to artifacts.

Our goal in this section was to explore the efficacy of using gesture to draw attention to artifacts during a scientific presentation. Given that comparisons between the non-visible gesture (NG) and visible gesture (YG) conditions (across all three types of gestures - ImplicitPointArtifact, ExplicitPointArtifact, and ArtifactManip) result in mean fixation time increases (five of the six statistically significant), we have strong evidence that gesture visibility plays a key role in drawing attention to artifacts. This evidence is further supported by the fact that the one paired condition that is not statistically significant is in the ArtifactManip condition, where even the weakest condition (NGNH) is highly effective at drawing attention to the artifacts in the presentation.

10.1.4 Impacts on FacialFeature AOIs across gesture types

Two of the key face-to-face visual communication channels are gesture and facial expression/features. Although our study is primarily interested in the effects of gesture on communication, it cannot do so at the expense of ignoring facial expression. There are two key questions that need to be answered in terms of how facial features are used in

communication. First, do subjects attend to facial features during a presentation, and if so how and when. Secondly, we need to provide a better of understanding of how attending to facial features impacts the attention to gestural communication and in particular how facial features impact attention to artifacts in scientific communication. We therefore look at the effect of facial feature visibility on both FacialFeature AOIs and on the three main gesture AOI types (ImplicitPointArtifact, ExplicitPointArtifact, and ArtifactManip).

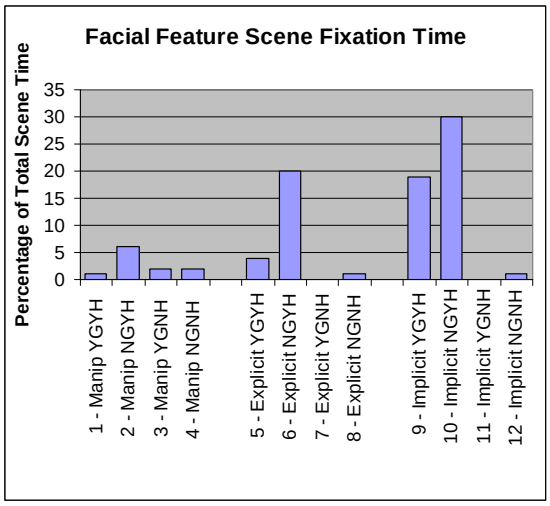


Figure 62: Percentage of fixation time for FacialFeature AOIs

10.1.4.1 Facial feature AOIs and gesture type

Unlike the comparison across gesture AOIs made in Section 10.1.3.1, FacialFeature AOIs only receive significant attention in the visible facial expression conditions. This is not surprising, as most facial feature AOIs are not collocated with artifacts on the screen. Therefore, when facial features are not visible, there is typically nothing of interest to attend to in the area of the FacialFeature AOIs. The percentage of total scene time for FacialFeature AOI fixations can be seen in Figure 62 (note that the FacialFeature AOI percentages displayed in Figure 62 are the same as those displayed in Figure 58). The percentage of total scene time spent in FacialFeature AOIs in the non-visible facial feature conditions are almost negligible, with 1% and 0% in the implicit artifact NGNH and YGNH conditions, 1% and 0% in the explicit artifact NGNH and YGNH conditions, and 2% and 2% in the NGNH and YGNH artifact manipulation conditions (Columns 4, 3, 8, 7, 12, and 11 in Figure 62).

Result 18 FacialFeature AOIs are attended to the most frequently when facial features are visible and gestures are not visible.

In addition, FacialFeature AOIs are not attended to consistently when facial features are visible. In particular, for all artifact gesture interaction types (manipulation, explicit artifact, and implicit artifact) the NGYH condition has the highest percentage of FacialFeature AOI attention.

Result 19 FacialFeature AOIs are attended to the least frequently in the YGYH condition during artifact manipulation scenes.

Interestingly, for both the artifact manipulation and explicit artifact gesture conditions where both gesture and facial features are visible (YGYH), FacialFeature AOI fixations are quite low. The artifact manipulation YGYH condition has the lowest FacialFeature AOI fixation percentage of all of the artifact manipulation conditions.

FacialFeature AOIs	NH	YH	Mean Difference	p-value	Test
ImplicitPointArtifact Scenes	NGNH	NGYH	15626	0.00001	Tamhane
	YGNH	YGYH	9816	0.0001	Tamhane
ExplicitPointArtifact Scenes	NGNH	NGYH	3017	0.002	Tamhane
	YGNH	YGYH	526	0.301	Tamhane
ArtifactManip Scenes	NGNH	NGYH	996	0.043	Tamhane
	YGNH	YGYH	-129	0.775	Tamhane

Table 9: Pair-wise comparisons of FacialFeature AOIs (varying H, constant G).

10.1.4.2 Facial feature AOIs and facial expression visibility

Result 20 Facial expression visibility significantly increases the attention drawn to FacialFeature AOIs for all but two of the pair wise comparisons when gesture visibility is kept constant.

When considering comparisons between facial feature visibility (NH to YH) across the gesture conditions (either YG or NG), in most cases the condition with the facial feature visibility has a significantly ($p < 0.05$) larger mean FacialFeature AOI fixation time (Table 9). Interestingly, this is not always the case. For explicit artifact and artifact manipulation scenes, the YGYH condition is not significantly different than the YGNH condition. In fact, in the artifact manipulation scenes, the FacialFeature AOI mean is lower (although not significantly) in the YGYH condition than in any other condition, including the NGNH condition which has no visual human communication cues at all.

FacialFeature AOIs	NG	YG	Mean Difference	p-value	Test
ImplicitPointArtifact Scenes	NGYH	YGYH	-6021	0.018	Tamhane
ExplicitPointArtifact Scenes	NGYH	YGYH	-2536	0.004	Tamhane
ArtifactManip Scenes	NGYH	YGYH	-1239	0.013	Tamhane

Table 10: Pair-wise comparisons of FacialFeature AOIs (varying G, constant H).

10.1.4.3 Facial feature AOIs and gesture visibility

Result 21 When facial features are visible, gesture visibility significantly reduces the mean fixation time in FacialFeature AOIs.

In all cases where facial features are visible (YH), the conditions without gesture visibility (the NGYH condition) have significantly higher mean FacialFeature fixation times than the non-visible gesture condition. The comparison statistics (using the Tamhane post-hoc mean difference test) across the YH conditions for each gesture type are shown in Table 10. Note that the mean difference is negative, indicating that the YG condition has a lower mean FacialFeature fixation time. These trends can also be seen in Figure 58, with the comparisons in Table 10 represented by Column 2 and 1, Columns 6 and 5, and Columns 10 and 9. This implies that there is strong evidence that gesture visibility reduces the mean fixation time in FacialFeature AOIs.

Result 22 Gesture visibility significantly increases the attention to artifacts while at the same gesture visibility significantly lowers the attention paid to facial features.

If at the same time we revisit the comparisons across these same conditions for the gesture based artifact AOI statistics (ImplicitPointArtifact NGYH and YGYH, ExplicitPointArtifact NGYH and YGYH, and ArtifactManip NGYH and YGYH comparisons from Table 7, Section 10.1.3.2), we see a statistically significant increase in the mean artifact fixation time across these conditions. Combining these two results, we see that in the YGYH condition there is significantly less FacialFeature fixation time and significantly more artifact (ImplicitPointArtifact, ExplicitPointArtifact, and ArtifactManip) fixation time than in the NGYH. A natural way of interpreting these results is that having visible gestures helps draw attention to the artifacts but it does so at the expense of lowering the attention paid to facial features. That is, the gesture and facial feature conditions are competing for "fixation time" and that artifacts attention tends to win the battle (we get more artifact attention and less facial feature attention).

It is worth pointing out that the same does not appear to be true when we compare across facial feature visibility while keeping the gesture condition constant (see Section 10.1.3.3 for a more detailed description). Only in the implicit artifact scenes is there a negative impact on artifact fixations (not statistically significant). For the explicit artifact and artifact manipulation scene types, there is either no significant decrease or a significant increase (ArtifactManip YGNH/YGYH) in the mean artifact fixation time.

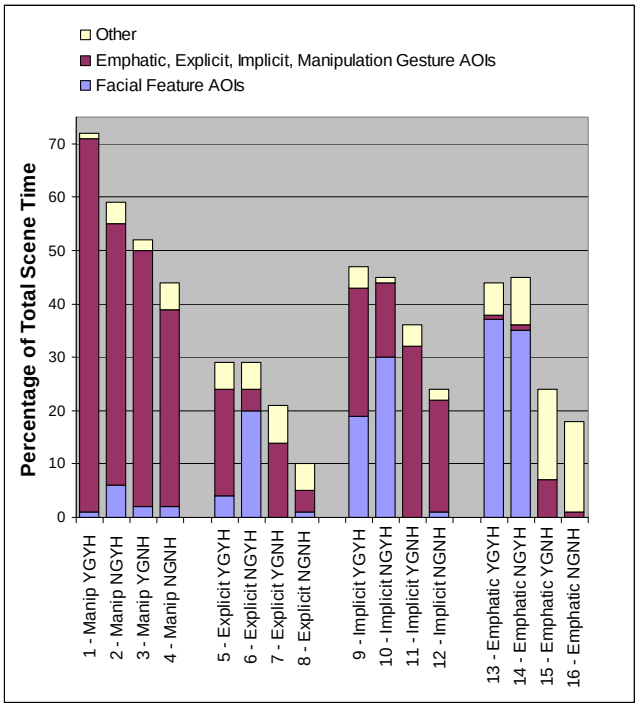


Figure 63: Fixation time for all AOI types in artifact related scenes.

10.1.5 Impacts on total AOI fixation time across gesture types

The total fixation time in AOIs across our conditions is also important to consider. We are interested in determining whether or not our experimental intervention increases the overall fixation time within all AOIs across the different gesture types. Our previous analyses show that both gesture and facial expression visibility have an impact on the fixation times in artifact and facial feature AOIs. An interesting question to ask is does the overall AOI fixation time change when facial expression or gesture visibility changes? Our previous results show that gesture and facial feature visibility impacts the attention for specific AOI types. If these changes in AOI attention occur without an increase in total AOI fixation time then our visibility conditions imply that there is a "competition" for attention between our AOIs. That is, an increase in facial feature attention would often result in a decrease in artifact attention (and vice versa). Alternately, if total AOI fixation time increases when gesture and facial features are visible (our visibility conditions increase overall attention time) then there may not be competition for attention between our AOI types. It is this question we consider below.

In considering the overall AOI fixation time (fixation in all AOI types), there is a clear pattern across all four gesture scene types (see Figure 63, reproduced from Figure 58 for clarity). The NGNH condition has the lowest percentage of total scene time spent in fixations (Columns 4, 8, 12, and 16 in Figure 63), followed by the YGNH condition (Columns 3, 7, 11, and 15), with the NGYH and YGYH conditions having the highest percentage of AOI fixations.

Total AOI Fixations	NG	YG	Mean Difference	p-value	Test
ImplicitPointArtifact Scenes	NGNH	YGNH	6552	0.024	Tukey
	NGYH	YGYH	765	0.986	Tukey
ExplicitPointArtifact Scenes	NGNH	YGNH	1708	0.032	Tukey
	NGYH	YGYH	-59	1	Tukey
ArtifactManip Scenes	NGNH	YGNH	1452	0.688	Tukey
	NGYH	YGYH	2486	0.276	Tukey

Table 11: Pair-wise comparisons of total AOI fixations (varying G, constant H).

10.1.5.1 Total AOI fixation time and gesture visibility

Result 23 Gesture visibility has a significant impact on total AOI fixation time in only a small number (two of the six) of pair wise comparisons.

When comparing the mean AOI times across the various conditions, few of the individual pairs show a significant difference. In particular, when keeping facial feature visibility constant (constant H) and considering different gesture visibility (varying G), only the implicit and explicit artifact gesture NGNH/YGNH comparisons show a significant increase in AOI fixations (see Table 11).

AOI Type	Scene Type	NG	YG	Mean Difference	p-value	Test
Total	ImplicitPointArtifact	NGYH	YGYH	765	0.986	Tukey
	ExplicitPointArtifact	NGYH	YGYH	-59	1	Tukey
	ArtifactManip	NGYH	YGYH	2486	0.276	Tukey
FacialFeature	ImplicitPointArtifact	NGYH	YGYH	-6021	0.018	Tamhane
	ExplicitPointArtifact	NGYH	YGYH	-2536	0.004	Tamhane
	ArtifactManip	NGYH	YGYH	-1239	0.013	Tamhane
Artifact	ImplicitPointArtifact	NGYH	YGYH	5487	0.012	Tukey
	ExplicitPointArtifact	NGYH	YGYH	1779	0.0005	Tukey
	ArtifactManip	NGYH	YGYH	4003	0.014	Tukey

Table 12: Comparisons of NGYH and YGYH across AOI and gesture types.

Result 24 There is no evidence that adding a gesture communication channel to the NGYH condition increases the overall AOI fixation time.

Result 25 Adding a gesture communication channel to the NGYH condition significantly increases the fixation time in artifact AOIs (ImplicitPointArtifact, ExplicitPointArtifact, and ArtifactManip AOIs).

Result 26 Adding a gesture communication channel to the NGYH condition significantly decreases the fixation time in FacialFeature AOIs.

Result 27 Since there is no evidence that the total AOI fixation increases when adding a gesture communication channel to the NGYH condition, the increase in attention to artifact AOIs comes at a cost of reduced attention to FacialFeature AOIs.

A deeper analysis of gesture visibility is important to consider. From Section 10.1.3.2 we know that the YGYH condition has a significantly higher mean artifact fixation time than the NGYH condition across all gesture types (the Artifact rows in Table 12). Since adding a gesture communication channel to the NGYH condition does not result in a significant total fixation time increase (the Total rows in Table 12), the statistically significant fixation time added to the artifact AOIs across these conditions implies that fixation time in other AOIs will be lower. From Section 10.1.4.3 we know that there is a significant decrease in mean FacialFeature fixation time across these conditions (the FacialFeature rows in Table 12). Our results therefore suggest that at least some of the increase in the artifact fixation time results from a decrease in the fixation time in FacialFeature AOIs.

Result 28 The effect of adding a gestural channel to the NGNH condition results in significant increases in artifact and total fixations for both implicit and explicit artifact scenes and no significant increase in artifact and total fixations for artifact manipulation scenes.

Note that this is not the case when adding a gestural channel to the NGNH condition. Comparing NGNH to YGNH conditions, we see that in the implicit and explicit artifact scenes both the artifact AOI fixations (NGNH/YGNH rows in Table 7) and the total AOI fixations increase significantly (NGNH/YGNH rows in Table 11). In the artifact manipulation scenes, neither the artifact manipulation AOI fixations nor the total AOI fixations significantly change. There is no significant change to FacialFeature AOIs, and indeed, because there is no facial expression visibility in both conditions there are almost no FacialFeature AOI fixations at all.

Total AOIs	NH	YH	Mean Difference	p-value	Test
ImplicitPointArtifact	NGNH	NGYH	11376	0.0001	Tukey
	YGNH	YGYH	5589	0.058	Tukey
ExplicitPointArtifact	NGNH	NGYH	3038	0.0001	Tukey
	YGNH	YGYH	1272	0.139	Tukey
ArtifacManip	NGNH	NGYH	2830	0.195	Tukey
	YGNH	YGYH	3864	0.024	Tukey

Table 13: Pair-wise comparisons of total AOI fixations (varying H, constant G).

10.1.5.2 Total AOI fixation time and facial expression visibility

Result 29 Facial feature visibility increases the total AOI fixation time in some pair-wise comparisons, but this increase is not consistent across gesture visibility (visible/not visible) or gesture types (implicit/explicit/manipulation).

When keeping gesture visibility constant (constant G) and varying facial feature visibility (varying H), only the implicit and explicit pointing NGNH/NGYH comparisons and the artifact manipulation YGNH/YGYH comparison result in significant differences (Table 13). The implicit pointing YGNH/YGYH comparison is also moderately significant ($p = 0.058$). This implies that facial feature visibility can help in increasing the amount of time fixated on AOIs, but this increase does not occur in all conditions.

AOI Type	Scene Type	NH	YH	Mean Difference	p-value	Test
Total	ImplicitPointArtifact	YGNH	YGYH	5589	0.058	Tukey
	ExplicitPointArtifact	YGNH	YGYH	1272	0.139	Tukey
	ArtifactManip	YGNH	YGYH	3864	0.024	Tukey
FacialFeature	ImplicitPointArtifact	YGNH	YGYH	9816	0.0001	Tamhane
	ExplicitPointArtifact	YGNH	YGYH	526	0.301	Tamhane
	ArtifactManip	YGNH	YGYH	-129	0.775	Tamhane
Artifact	ImplicitPointArtifact	YGNH	YGYH	-4030	0.068	Tukey
	ExplicitPointArtifact	YGNH	YGYH	324	0.821	Tukey
	ArtifactManip	YGNH	YGYH	4186	0.006	Tukey

Table 14: Comparisons of YGNH and YGYH across AOI and gesture types.

Result 30 In implicit artifact scenes, facial feature visibility significantly increases FacialFeature AOI fixations.

Result 31 In artifact manipulation scenes, facial feature visibility significantly increases the overall fixation time and the ArtifactManip AOI fixation time.

Unlike the consistent results we see when adding a gestural channel to the NGYH condition (comparing NGYH/YGYH), the results when adding a facial expression channel to the YGNH condition (comparing YGNH/YGYH) are not as clear cut (Table 14). For implicit artifact scenes the facial expression channel significantly increases the

FacialFeature AOI fixation time ($p = 0.0001$) but only moderately increases the total AOI fixation times ($p = 0.058$). Interestingly, it also moderately increases the ImplicitPointArtifact AOI fixation times ($p = 0.068$). For explicit artifact scenes, the facial expression channel has no significant impact on any AOI fixation times, including no impact on the FacialFeature AOI fixation time. The most intriguing results from this analysis is that the facial expression visibility increases the ArtifactManip AOI fixation time ($p = 0.006$), increases the total fixation time ($p = 0.024$), but there is no evidence that it increases the FacialFeature AOI fixation time ($p = 0.778$).

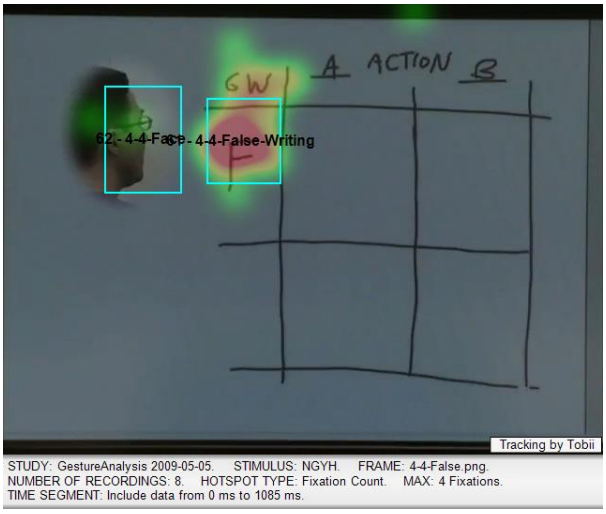


Figure 64: Facial feature acting as a pointing mechanism.

Result 32 Facial features, and in particular the spatial location and gaze direction of the face, can act as a pointing gesture.

These results are counter intuitive. Not only did adding facial feature visibility fail to increase the FacialFeature time as one might expect, it also did not decrease the fixation time in the ArtifactManip AOIs. In fact, it did exactly the opposite, causing a significant increase in the amount of attention paid to the ArtifactManip AOIs. One possible explanation for the increase in ArtifactManip AOI attention when facial expression is visible is that in some ways the facial features can be construed as a crude pointing device. For example, when the presenter states "… lets put F for the future where it turns out to be false …" he is looking at the top left of the Pascal's Wager diagram where the writing is taking place (see Figure 64). Note that both the spatial location of the face as well as the gaze direction of the presenter provides spatial cues as to which artifact is being discussed and/or manipulated. Although not a direct pointing gesture, our results

suggest that such facial feature cues, combined with the natural attention that artifact manipulation draws by itself, are what cause this increase in mean artifact fixation time in the artifact manipulation YGYH condition.

10.1.6 Effectiveness of gesture types

We consider three types of artifact gesture in this study. Recall that an implicit artifact gesture is used to reinforce the communication taking place. That is, the communication can be understood without the gesture because the utterance has implicit information about the referent artifact. In an explicit artifact gesture/utterance pair, the gesture carries with it critical information ("this box"), and without the explicit gesture the communication cannot be understood in its entirety. The deictic nature of the gesture implies that subjects need to pay more attention to the gesture in order to ascertain the correct meaning of the utterance. Artifact manipulation gestures are those that are made while manipulating an artifact (writing or marking up the diagram). During the manipulation process, not only does the gesture draw attention to the artifact but the artifact is changing dynamically during the gesture.

	FacialFeature (%)	Artifact (%)	Ratio
ArtifactManip YGYH	1	70	70.0:1
ExplicitPointArtifact YGYH	4	20	5.0:1
ImplicitPointArtifact YGYH	19	24	1.3:1

Table 15: Ratio of Artifact to FacialFeature percentages for the YGYH condition.

Result 33 When both facial features and gesture are visible (YGYH), artifact manipulation gestures are the most effective at focusing attention on artifacts, followed by explicit artifact gestures. Implicit artifact gestures are the least effective at drawing attention to artifacts.

It is interesting to consider measures of effectiveness in terms of attracting attention to artifacts. We do this by considering the ratio of the percentages of artifact AOIs to FacialFeature AOIs when both are visible in the scene (the YGYH condition). These ratios can be seen in Table 15. The Artifact to FacialFeature ratio in the YGYH condition is by far the highest in the artifact manipulation scenes at 70:1, with the explicit artifact gesture scenes second with a 5:1 ration, and the implicit artifact gesture scenes with the lowest ration of 1.26:1. This can also be seen pictorially by considering the relative sizes of the red and blue bars in Column 1, Column 5, and Column 9 in Figure 63 (p. 249). Clearly, the ArtifactManip gestures are extremely effective at drawing attention to

artifacts and our results suggest that this is done at the expense of attention to the FacialFeature AOIs. Explicit artifact gestures also exhibit a high effectiveness of drawing attention to artifacts, again at the expense of attention to FacialFeature AOIs. Implicit artifact gestures, although still effective at drawing attention to artifacts, do not focus that attention like artifact manipulation and explicit artifact gestures do.

Result 34 Artifact manipulation and implicit artifact gesture/utterance pairs are effective at drawing attention to artifacts, even when the gestures are not visible (NGNH).

Result 35 Explicit artifact gesture utterance pairs are not effective at drawing attention to artifacts when gesture is not visible, with little difference between the two conditions where there is no gesture visibility (NGNH, NGYH).

When considering the three artifact gesture scene types, it is interesting to note that the NGNH condition has a significant amount of fixation time within the artifact AOIs, even though there is no gestural cue available to the subjects in those conditions. This is primarily due to the fact that in the NGNH condition, the only things to fixate on are the artifacts on the whiteboard. These fixations are either incidental fixations, fixations due to action (writing or circling an artifact), or fixations that are directed by utterances made by the presenter.

In both the NGNH implicit artifact and artifact manipulation scenes, there are other cues that the presenter uses to draw attention to the artifacts. AOI fixations will sometimes result from the subjects looking at artifacts that are referred to by the speaker in the utterance during an implicit artifact gesture. For example, when the presenter says “if global warming is true” and points at the T in the diagram, although the gesture itself will not be seen in the NGNH condition it is possible for a participant to determine that the T is the referent artifact through the utterance alone. Similarly, when the presenter is writing in the NGNH condition, although the writing action (the writing gesture) is not visible, the actual artifact does appear in the diagram as the writing action takes place. The change in the diagram itself will sometimes draw fixations to the artifact in question. Figure 63 (p. 249) shows that this may indeed be the case, with the NGNH implicit artifact gesture and artifact manipulation scenes having a larger percentage of gesture fixations (Column 4 and 12) than the NGNH explicit artifact scenes (Column 8).

Result 36 When only one of the two human communication channels is visible, the facial feature channel generates more AOI fixations than the gesture communication channel.

Finally, we consider the comparisons from NGNH to both the YGNH and NGYH conditions across all scene types. In effect, we are comparing the addition of either a gestural communication channel (YG) or a facial feature communication channel (YH) to a condition that has no visual human communication channels to begin with (NGNH). In all cases, the NGYH condition has a larger increase in the percentage of total scene time than YGNH condition within the same type of gesture scene. This implies that adding facial features communication channel attracts more attention than the gestural communication channel.

It is not clear whether this is due to the fundamental nature of the two types of communication channels or whether it is due to the fidelity and quality of the communication channels presented in the study. In particular, the fidelity of the pointer used to communicate gestural information in the YGNH condition may have contributed to this difference. The facial feature condition is quite realistic, and may have attracted more attention because of its novelty (a disembodied head moving across the screen). At the same time, we were unable to create a realistic pointer that equalled the fidelity of the facial feature condition. The traditional computer-style cartoon hand pointer used, although very dynamic in its motion, is of relatively low visual fidelity compared to the facial feature cues provided. Thus, this low fidelity may have had an impact on the attention paid to artifact AOIs. The fidelity of the visual gestural channel is certainly something that would be worth investigating in more detail.

10.2 Measures of Task

Our analysis above explores measures of the communication process. What communication channels are attended to and used to decode information? What are the impacts of our experimental interventions on attention? Ultimately, the goal of a research presentation is to impart understanding on the listeners. Clearly, it is necessary to understand how information is communicated. It is equally as important to understand how the decoding of information translates into understanding. It is this area that our measures of task explore.

10.2.1 Impacts of facial feature and gesture visibility on questionnaire responses

The post-study questionnaire is designed to determine whether our experimental interventions affect the understanding of the topic of the presentation. Recall that the questionnaire was designed to test three main aspects of the subject's understanding:

- Understanding about the artifact: We test our subject's understanding of the structural nature of the Pascal's Wager diagram and in particular the roles the rows and columns play in the presentation (Question 2 and 3).
- Understanding about information: We test our subject's understanding of the information presented using the Pascal's Wager diagram and in particular the recollection of specific facts that were presented during the presentation (Question 4 and 5).
- Understanding about the argument: We test our subject's understanding of the argument being made by the presenter and in particular the subject's recollection of several key facts that the presenter used to make his argument (Question 6 and Question 7).

Result 37 Gesture and facial feature visibility did not have an effect on the subject's understanding of the presentation ($\alpha = 0.05$)

The results from the analysis presented in Appendix 15.14.2.5 reveals that our experimental intervention did not have a significant impact on our subject's understanding of the presentation (at an $\alpha = 0.05$ level). Although there are no statistically significant results at the $\alpha = 0.05$ level, there are some conditions in which there are moderately significant results (at the $\alpha = 0.1$ level). We consider these results below.

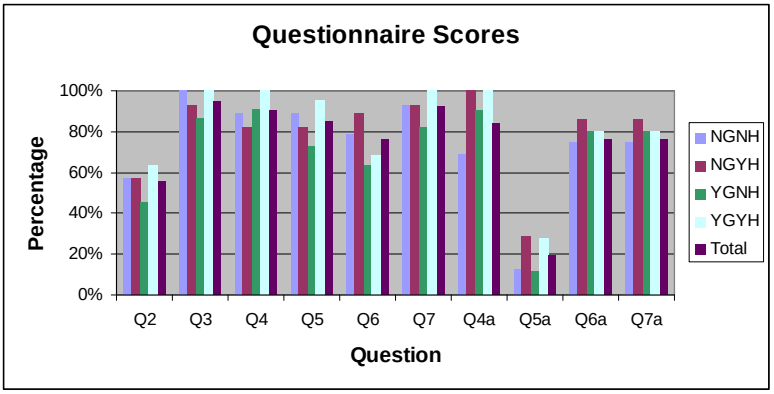


Figure 65: Questionnaire scores for all questions.

In looking at the scores in detail, it is worth noting several details. Question 2 focuses on whether or not the subjects understood that the top row in the Pascal's Wager diagram represented the question of whether or not humans are the cause of global warming (true or false). Many subjects answered the question stating that the row represented whether or not global warming was really occurring or not. Subjects were awarded one mark for this answer and two marks for the correct answer (whether humans are the cause of global warming). The average overall score across all conditions was 1.12 (56%), implying that most subjects got this question partially wrong (40 of the 50 subjects scored one mark out of two, with only eight getting it correct). For all other questions, the average across all conditions ranged from 76% (Question 6) to 95% (Question 3). See Figure 65 for the overall questionnaire score percentages for each question in each condition and the percentage for each question across all conditions. We see similar results for Question 5a. Question 5a is marked out of a total of five marks, with an average score of 1 (20%). Eighteen of the subjects scored either 0 or 1. Again, for all other secondary questions (Q4a, Q6a, and Q7a) participants scored relatively well, with average scores of 84%, 76%, and 76% respectively.

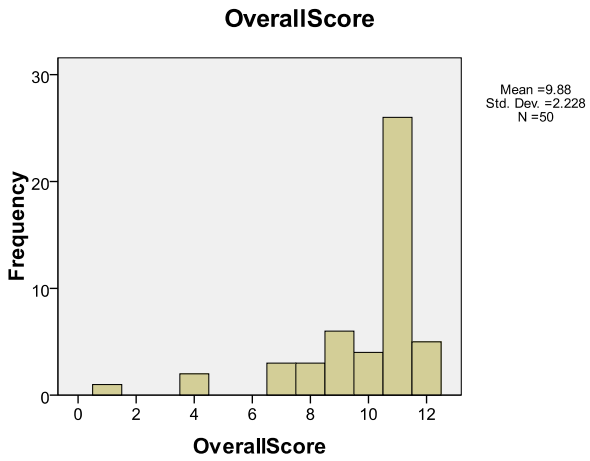


Figure 66: Histogram of the Overall scores (Q2 – Q7).

This implies that most of the questions were answered relatively well and that subjects across all conditions had a basic understanding of most of the key points of the presentation. This includes understanding of the structure of the diagram used, the information presented in the diagram, and the concepts used in the argument made by the presenter. Overall, the average score on the basic questions was 9.8 (out of a possible 12), with a large grouping of subjects at the 11 out of 12 level (see Figure 66).

We first consider the pair-wise analysis of the YGYH and YGNH conditions (see Section 15.14.2.2 for details). These two conditions were investigated further because these conditions consistently received either the highest (YGYH) or the lowest (YGNH) questionnaire scores. The YGYH condition had the highest average score on five of the six questions, as well as on the overall score. This is not surprising, as it is the richest visual condition in the study. The YGNH conditions had the lowest average score on five of the six questions as well as in the overall score. When compared, the means ranks of these two conditions are moderately significantly different on the Overall, Question 5, and Question 7 scores ($0.05 < p < 0.1$).

Question 6 shows a similar significance level when comparing the conditions where both gesture and facial feature visibility change (across the NGYH and YGNH conditions), with the NGYH condition mean score being the highest and the YGNH condition mean score being the lowest a ($p = 0.089$). It is also worth noting that Question 6 is also the only question where the YGYH condition is not the highest scoring condition. It is interesting to note that with the exception of Question 2 (discussed above),

Question 6 is the question that is the most poorly answered (with an average score of 76%). Although it is difficult to conclude anything from this analysis, it suggests that on Question 6 gesture visibility may have a negative impact on questionnaire score (adding gesture visibility to the NGYH condition results in a lower score). Since the statistics are only marginally significant, we can only infer that this area requires further study.

10.2.2 Impacts on facial feature and gesture visibility on extended questionnaire responses

Result 38 There is a trend towards facial feature visibility having an impact on the Overall, Question 4a, and Question 5a scores but this trend is not significant at a level of $\alpha = 0.05$.

Result 39 There are no main effects for gesture visibility.

When considering the supplemental questions (Question 4a – Question 7a) we see similar, but slightly more robust results. The Kruskal-Wallis test suggests that there are moderate statistical effects for differences in the mean ranks of Question 4a ($p = 0.063$) and the two-way ANOVA performed on Question 5a and the Overall scores shows a moderately significant effect for facial feature visibility ($p = 0.081$ and $p = 0.068$ respectively). Note that in the analysis of Questions 2 through 7 (Section 10.2.1) none of the one-way Kruskal-Wallis mean rank or one-way ANOVA tests showed any statistical significance. It was only the pair-wise Mann-Whitney U tests in a small number of pairings that showed moderately significant results. Although the results for Question 4a through 7a are still only moderately significant, the tests used are the more robust ANOVA tests (two-way ANOVA and one-way mean rank Kruskal-Wallis ANOVA) and result in moderately significant main effects across conditions. We can therefore consider these moderately significant results with a higher degree of confidence than the results for Questions 2 through 7.

Result 40 Facial feature visibility has a significant impact ($p = 0.036$) on mean rank score on Question 4a when no gesture is visible (NGNH/NGYH).

Given that we have a moderately significant result ($p = 0.063$) on Question 4a, we perform pair-wise Mann-Whitney U tests to determine which conditions are significantly different. The NGYH condition is a significantly higher mean rank score than the NGNH condition ($p = 0.036$) while the YGYH condition has a moderately significant higher mean rank score than the NGNH condition ($p = 0.071$).

Recall that the maximum scores for Question 4a, 5a, 6a, and 7a are two, five, two, and two respectively. Question 5a is marked out of five because the subject is asked to list all of the "global disasters" that are presented in the lower right quadrant of the diagram (there are five). The average scores for the four questions are 84%, 20%, 76%, and 76% respectively (see Figure 66). Like Question 3 through Question 7, Question 4a, 6a, and 7a have relatively high average scores. The subjects answered these questions fairly accurately. Question 5a, on the other hand, has an average score of one out of five. Subjects in general were unable to answer this question successfully, with 18 of the 25 subjects scoring zero or one. Although there is a moderately significant main effect of facial feature visibility on the mean scores for Question 5a ($0.05 < p < 0.1$), our analysis of our experimental interventions does not demonstrate a clear reason for the poor performance on Question 5a.

Result 41 On the scenes relevant to Question 5a there is a significant increase in the fixation time spent in artifact AOIs (p = 0.05) between the NGYH and YGYH conditions.

In order to investigate this question in more detail we consider the AOI fixations in the scenes that involve the answers to Question 5a (see Section 15.14.2.4 for the detailed analysis). An initial analysis indicates that gaze fixations are fairly typical in these scenes. There is a significant interaction effect between gesture and facial feature visibility on fixations in artifact AOIs in these scenes ($p = 0.039$). When facial features are visible there is a significant difference across the gesture visibility conditions, with significantly more artifact fixation time when gestures are visible (Tukey HSD, $p = 0.05$). When facial features are not visible, there is little evidence that gesture visibility has an impact on artifact fixations (Tukey HSD, $p = 0.990$).

Result 42 On the scenes relevant to Question 5a facial feature visibility has a significant effect on the mean total time spent in AOIs ($p = 2.46 \times 10^{-5}$).

We also demonstrate a trend toward a significant main effect of facial feature visibility on mean artifact fixation time in the artifact creation scenes ($p = 0.096$) and a highly significant effect of facial feature visibility on the overall mean fixation time in all AOIs ($p = 2.46 \times 10^{-5}$).

Thus, as in most of the scenes that involve artifacts, both gesture and facial feature are effective at drawing attention to the artifacts required to answer Question 5a. In

particular, facial feature visibility figures prominently in all of the statistically significant analyses listed above, including having an effect on both the effectiveness of gesture and the score on Question 5a. Despite these results, these links are somewhat tenuous. The effect on score is only moderately significant and the direct effects of gesture visibility on score are not present.

Result 43 Scores on Question 5a could have been confounded by the use of other artifacts to highlight other plausible, yet incorrect answers to Question 5a.

We are therefore forced to consider that the key contributing factor to the low scores on Question 5a are outside the controls of the experiment. It may in fact be the case that the question was simply too difficult. The video is quite long, the artifact manipulation that reveals the correct answer to the question occurs early in the video (three minutes into the ten minute video), and the participants are presented with a lot of information throughout the video. Even though the question is quite explicit and the creation of the artifact and discussion around it is quite important to the presenter's argument, it appears that subjects simply could not remember the five items listed in the diagram.

One possible confounding factor is the fact that during the video the presenter discusses many global disasters that could result from global warming. For example, in the "Cans" section of the video (see Section 9.2 and Figure 52 for details), the presenter lists many other "disasters" that could occur as a result of global warming. Recall that Question 5a asks explicitly for only those "disasters" that are listed in the Pascal's Wager diagram. Thus, participants were faced with the task of remembering five out of approximately twenty alternatives. In looking at the actual responses, this appears to be at least part of the problem. Some participants answered with five disasters but listed disasters that were in the "Cans" section of the video but not in the diagram. Some participants failed to list more than one or two disasters, either not realizing that there were more in the diagram, realizing there were more and not being able to remember, or ignoring the question's specific directions to list them all. There are specific examples of each of these cases in subject's questionnaire responses.

11 Gesture Study: Summary

One of the key objectives of the research presented in this dissertation is to help inform the design process for the creation of new collaboration tools. One key factor in reaching this objective is to understand how scientific collaborators process and decode the meaning and intent that is communicated during scientific presentations. The goal of our gesture study is to form an understanding of how researchers process the myriad of visual cues, including gestures, which are necessary to communicate a clear and concise understanding of a complex scientific problem. It is not enough to know that researchers use gesture extensively (Chapter 7). Nor is it enough to know that current collaboration tools do not communicate gesture effectively (Chapter 7). These two facts cause a problem if, and only if, gesture is important to the effective understanding of the intended communication. We do not need to design collaboration tools that transmit gestural interaction if researchers do not attend to such gestures and/or such gestural communication does not add to the efficacy of the communicated message.

The gesture study presented in Chapters 8, 9, and 10 helps us to *"Develop a deep understanding of how scientific researchers interact with digital artifacts when they collaborate."* (Objective 2:). In particular, it focuses on exploring the following research question in detail: *"What communication channels are used to decode information during artifact-centric collaboration?"* This leads to a number of other important questions revolving around the two main visual mechanism humans use to communicate – gesture and facial features. The first set of questions revolves around gesture. Does pointing at an artifact draw attention to that artifact? Does pointing at an artifact help an observer understand the artifact better? Assuming the artifact pointed at plays an important role in a discussion, does pointing at the artifact help an observer understand the role the artifact plays in the discussion better?

The second set of questions revolves around facial features. Do observers pay attention to facial features during a discussion? What do observers attend to when facial features, artifacts, and gestures are all involved in a discussion? Do facial features have an impact on the understanding observers have about artifacts when they are used as part of a

discussion? And lastly, how do facial features and gesture interact when carrying out a complex discussion that involves artifacts?

These questions are captured in the hypotheses originally presented in Section 8.2. We study the results from our gesture study (as presented in Chapter 9 and Chapter 10) in the context of these hypotheses below. We first consider the impact of our experimental interventions on the hypotheses that involve our process measures (attention, as measured by gaze fixation) (Section 11.1). We then consider the impact of our interventions on the hypotheses that involve our task measures (understanding, as measured by questionnaire scores) (Section 11.2).

11.1 Impact of Experimental Interventions on Process Measures

11.1.1 Impacts of Gesture Visibility on Artifact Attention

Hypothesis 1: Researchers will attend to artifacts when they are used as part of a presentation.

Hypothesis 2: Researchers will attend to artifacts more frequently when gesture is used to draw attention to an artifact.

Our results from the study of global fixation phenomena are the building blocks we use throughout the remainder of this chapter. These results show that eye fixations (**Result 1**), and in particular fixations within the AOIs used in this study (**Result 2**), are effective at capturing our participant's attention. Our global analysis also shows that our study participants spend a large percentage of fixation time attending to artifacts that are manipulated as part of a scientific presentation (**Result 4**). In addition, our first explorations of artifact gestures indicate participants attend to artifacts that are referred to as part of a pointing gesture (**Result 5**).

Although our global phenomena analysis provides a basic analysis of attention, the core of our analysis revolves around our experimental interventions. Our analysis provides strong support for our hypotheses that participants attend to gesture, and gesture AOIs, across almost all gesture types and all conditions. The level of attention and in which conditions there is a significant difference depends on the type of gesture.

Gesture visibility has a significant impact on the level of attention paid to emphatic gesture AOIs, but the difference is only significant when facial features are not visible

(**Result 8**). Gesture visibility has almost no impact on fixation times in other emphatic gesture conditions, on facial feature fixations, and on total AOI fixation time (**Result 9**). Emphatic gestures are attended to less than any of the other gesture types (**Result 10**).

Gesture visibility has a significant effect on the attention paid to artifact AOIs for all the artifact gesture types (artifact manipulation, implicit artifact gestures, and implicit artifact gestures) and across all pairings of gesture visibility conditions with the exception of the NGNH/YGNH artifact manipulation condition (**Result 11**, **Result 16**). This is statistically the most consistent and the strongest result in this study (see Section 10.1.3.2 and Table 8). *These results clearly show that gesture visibility plays a critically important role in drawing attention to artifacts during scientific presentations.*

Scenes that contain artifact manipulation gestures have the largest percentage of artifact AOI fixation time across all conditions (**Result 13**). Implicit artifact AOIs have a higher percentage of artifact fixations than explicit artifact AOIs (**Result 15**). The explicit artifact AOIs in the no gesture visibility conditions have the lowest percentage of fixation times of all of the other conditions (**Result 14**). Given that explicit artifact gestures require gesture visibility (the deictic utterances depend on gesture to be communicative), this is not a surprising result. At the same time, it is an important result, as it implies that for explicit artifact gestures to effectively draw attention to artifacts the gesture must be visible. Recall that this was one of the questions that were raised in our ethnography (Chapter 7).

11.1.2 Impacts of Facial Feature Visibility on Artifact Attention

Hypothesis 4: Researchers will attend to facial expression when it is communicated as part of a presentation.

Hypothesis 5: Researchers will attend to the artifacts used in a presentation less when facial expression is visible as part of the presentation.

In considering the impacts of facial feature visibility on artifact attention, we first revisit our analysis of the study of global phenomena. All participants spent a significant amount of time attending to facial features when they were visible on the screen (**Result 3**). Our comparative analysis of the attention paid to facial features show that participants do not attend to facial expression as often when artifact manipulation (**Result 6**) and artifact pointing gestures (**Result 7**) are used as part of a scientific presentation. Although

this does not provide evidence for the hypotheses above, these results do suggest that artifact and facial feature fixation time may be related.

Our analysis of facial feature attention across our experimental conditions shows that facial feature visibility has a main effect on AOI attention for most AOI types and most artifact gesture scenes (**Result 12**). Facial feature AOIs are attended to the most frequently when facial features are visible and gesture is not visible (**Result 18**). When only one of the two visibility conditions is visible (either facial feature or gesture, but not both), facial feature visibility generates more AOI fixations than the gesture visibility across all three artifact gesture types (**Result 36**). Interestingly, in artifact manipulation scenes, facial feature AOIs are attended to the least frequently in the condition where both gesture and facial features are both visible (**Result 19**). Facial feature visibility increases facial feature AOI fixations for some, but not all gesture types (**Result 20**).

As we saw above, gesture visibility draws attention to artifact AOIs (**Result 11**, **Result 16)**. Our results also suggest that gesture visibility reduces the attention spent on facial feature AOIs (**Result 21**, **Result 22**). The inverse is not always true. Facial feature visibility sometimes decreases, sometimes has no effect, and sometimes *increases* the amount of attention participants attend to artifact AOIs.

Although these results demonstrate that in some conditions there is strong support for Hypothesis 4: and Hypothesis 5: , this support does not exist for all conditions. Indeed, there is strong support that in some conditions the opposite is true. That is, facial feature visibility increases the amount of attention paid to artifacts. This result is important, as it means adding the facial feature visibility to a collaboration tool will not necessarily have a negative impact on the attention paid to artifacts. In fact, it sometimes helps.

11.1.3 Interactions between Gesture and Facial Feature Visibility

In order to definitively determine the impact that facial feature visibility has on artifact attention, it is necessary to consider the total time spent within all AOIs in a scene. That is, if adding more communication channels simply means an increase in overall AOI fixation times and no negative impact on desirable AOI fixations (e.g. artifact fixations), then there is no reason not to add more channels. On the other hand, if adding a communication channel does not increase overall AOI fixation time and thereby reduces fixations on important AOIs (e.g. artifacts) then adding channels may not be wise.

Our results above show that in many conditions adding facial feature visibility does not have a dramatic negative impact on artifact AOI fixations. Our analysis of total AOI fixation time indicates that both gesture and facial feature visibility tend to increase the overall AOI fixation time. This is not surprising, since our AOIs revolve around facial feature and artifact AOIs. What is interesting is the fact that the increase is not consistent across our experimental conditions (facial feature and gesture) or artifact interaction types (implicit, explicit, and manipulation). For example, gesture visibility has a significant impact on total AOI fixation time in only a small number of pair-wise conditions (**Result 23**). The same holds true for facial feature visibility (**Result 29**).

This interaction is quite complex. In the case of facial features, adding the facial feature visual channel tends to add more AOI fixation time and is one key reason why the fixation time on artifacts does not decrease. In those cases where comparisons are made between two conditions that differ in a visual channel (e.g. gestural channel) and there is no increase in the overall fixation time, it is inevitable that the AOI fixation time will be shared among the AOIs of the visual channels that are visible (**Result 27)**. That is, adding the visible gestural channel does not result in an increase in overall fixation time (**Result 24)** but it does result in an increase in artifact AOI fixations (**Result 25)** and a decrease in facial feature AOI fixations (**Result 26**). At the same time, adding a gestural channel to the NGNH condition of the implicit and explicit artifact scenes results in an increase in both the total AOI time and the artifact AOI time (**Result 28)**. In implicit artifact scenes, facial feature visibility increases facial feature, artifact, and total AOI times (**Result 30)**.

Of particular interest is the impact of facial feature visibility in the artifact manipulation scenes. When facial feature visibility conditions are compared to those without facial feature visibility, the visible facial feature conditions increase the overall AOI fixation time, have no significant effect on the fixation time in facial feature AOIs, and increase the artifact AOI fixation time (**Result 31**). In effect, our results show that facial features, and in particular the spatial location and gaze direction of the eyes, can act as a pointing style gesture (**Result 32**). This result is quite surprising, and indeed very encouraging. Not only does facial feature visibility typically not have a negative impact on artifact attention, it sometimes helps significantly.

Considering the ratio of artifact fixation time and facial feature fixation time provides us with a measure of how facial feature visibility impacts artifact attention in a given scene type. We perform these comparisons in the YGYH condition, where both facial features and gesture are visible. The ratio of artifact to facial feature fixation percentage clearly indicates that artifact manipulation interaction is the most effective at focusing attention on artifacts (70:1), followed by explicit artifact gestures (5:1), and implicit artifact gestures (1.3:1) (**Result 33**). Note that considering effectiveness in this manner gives us a different view than we had when considering percentage of time in fixation (**Result 13, Result 15**). Although implicit artifact interactions result in a higher percentage of total artifact fixation time than explicit artifact interaction (**Result 15**), by our ratio measure explicit artifact interaction has a higher *efficiency* in terms of focusing attention on artifacts (**Result 33**).

We conclude our analysis by considering the NGNH condition in two ways. First, we note that both artifact manipulation and implicit artifact gesture/utterance pairs are effective at drawing attention to artifacts, even when the gestures are not visible (**Result 34**). In the artifact manipulation interactions, the artifact changes and therefore naturally draws attention to the artifact. In implicit artifact interactions, recall that the utterance that accompanies the interaction contains enough information to at least partially imply the referent artifact. Thus, both artifact manipulation and implicit artifact interactions contain enough information to infer some knowledge about the artifacts being discussed. Note that this is not the case in explicit artifact interactions, and as a result there is very little artifact fixation in the NGNH explicit artifact condition. It is also worth noting that in the explicit artifact interaction conditions, where no gesture is visible, there is very little difference between artifact fixation times (**Result 35**).

11.2 Impact of Experimental Interventions on Task Measures

Before considering the impacts of gesture visibility on questionnaire scores, we consider the overall understanding obtained by our participants. Participants had a good overall understanding of the concepts addressed by the questionnaire, including the structure of the diagram used, the information presented in the diagram, and the concepts used in the argument made by the presenter. At the same time, there were two important

concepts that our questionnaire addressed that resulted in very poor results on Q2 (average of 56%) and Q5a (average of 20%).

11.2.1 Impacts of Gesture Visibility on Questionnaire Scores

Hypothesis 3: Researchers will have a better understanding about artifacts, how they are used, and the information they contain when gesture is used to refer to those artifacts during a presentation.

Our results show very little evidence that gesture visibility has an impact on understanding as measured by our questionnaire. Gesture visibility did not have a statistically significant impact on overall questionnaire score, on any of the individual question scores for Questions 2 – 7 (**Result 37**), and on any of the individual scores for Question 4a – 7a (**Result 39**). On Question 4a we see a moderately significant difference among the mean ranks scores ($p = 0.063$) with a moderately significant pair-wise increase between the no gesture and no facial feature visibility condition (NGNH) and the gesture and facial feature visible conditions (YGYH). There are no other pair-wise differences on this question. *Our results from this study provide little to no evidence that gesture visibility has a significant impact on questionnaire scores.*

11.2.2 Impacts of Facial Feature Visibility on Questionnaire Scores

Hypothesis 6: Researchers will have a poorer understanding about artifacts, how they are used, and the information they contain when facial expression is visible as part of the presentation.

Our results do show that facial feature visibility has an impact on questionnaire scores. Although facial feature visibility did not have a statistically significant impact on overall questionnaire scores or on any of the individual question scores for Questions 2 – 7 (**Result 37**), there are some indications of an effect across some conditions on some questions. In particular, in pair-wise comparisons on Question 5, Question 7, and Overall scores we see a moderately significant increase in score between the visible gesture and no visible facial feature condition (YGNH) and the visible gesture and visible facial feature conditions (YGYH). There is a trend ($0.05 > p > 0.1$) towards statistical significance of facial feature visibility on the overall, Question 5a and Question 4a scores (**Result 38**). On Question 4a there is statistical significance in pair-wise increases in

questionnaire score between the no gesture and no facial feature visibility condition (NGNH) and the no gesture and visible facial feature conditions (NGYH) (**Result 40**). *Of critical importance to note is that our results suggest that facial feature visibility increases, not decreases, the questionnaire score. Thus, not only is there no evidence to support our hypothesis, but we have relatively strong evidence that suggests facial feature visibility increases questionnaire scores.*

Although this is a surprising result, this finding has support from our results on process measures. Recall that in Section 10.1.5.2 we demonstrate that facial features can act as a pointing gesture and can in fact draw attention to artifacts (**Result 31**, **Result 32**). If this is indeed the case, then facial feature visibility having a positive impact on questionnaire score is not so surprising.

11.2.3 Exploring Question 5a

Since the scores on Question 5a are so poor (average score of 20%) and that this question shows some impact from our experimental conditions, we explored this question in significant detail. Recall that there is a trend towards facial feature visibility increasing questionnaire scores on Question 5a but this trend is not significant at the $\alpha = 0.05$ level (**Result 38**). We performed a detailed analysis of the scenes in which the artifacts that are required to answer Question 5a are referred to using artifact gestures. In these scenes, gesture visibility has a significant effect ($p = 0.05$) on fixation time in artifact AOIs, but this effect is only present when facial features are visible (in the YH conditions) (**Result 41**). There is a trend towards significance in the main effect of facial feature visibility on fixation time in artifact AOIs, but again this effect is not at the $\alpha = 0.05$ level ($p = 0.096$). There is a significant main effect of facial feature visibility on the total time spent in AOIs ($p = 2.46 \times 10^{-5}$) (**Result 42**). Thus, as in most of the scenes that involve artifacts, in some conditions gesture visibility is effective at drawing attention to the artifacts required to answer Question 5a. *Although it is worth noting that facial feature visibility is noticeable, if not statistically significant, in the analyses listed above, our results are inconclusive. More research needs to be performed in this area to gain a better understanding of these interactions.*

It is worth pointing out that there is some evidence that the correct answers on Question 5a may have been confounded by the fact that there were a number of other plausible, yet incorrect answers to the question described throughout the presentation. Participants were asked to list the set of five specific disasters that were listed in the diagram. Throughout the presentation a number of other disasters were discussed by the presenter. An analysis of questionnaire answers shows that many participants answered with real disasters discussed in the presentation but *not* the correct disasters specifically listed in the diagram (**Result 43**).

11.3 Discussion

The study presented in this Chapter is critical in providing a deep understanding of how scientists interact with data when they are collaborating (Section 1.2, Objective 2:). In particular, it provides new information that helps to answer the research question "*What communication channels are used to decode information during artifact-centric collaboration?*" We summarize the results regarding the research hypotheses for this study below.

Hypothesis 1: Researchers will attend to artifacts when they are used as part of a presentation.

Hypothesis 2: Researchers will attend to artifacts more frequently when gesture is used to draw attention to an artifact.

Our analysis shows that Hypothesis 1 and Hypothesis 2 are well supported by our results. Gesture visibility has a significant impact on the time spent attending to artifacts AOIs during a scientific presentation.

Hypothesis 4: Researchers will attend to facial expression when it is communicated as part of a presentation.

Hypothesis 5: Researchers will attend to the artifacts used in a presentation less when facial expression is visible as part of the presentation.

Support for Hypothesis 4 and Hypothesis 5 is not consistent across all conditions. Our results show that facial feature visibility sometimes decreases attention to artifacts as our hypotheses suggest, but this is not always the case. In some cases facial feature visibility increases the time spent attending to artifacts. Of particular interest is the fact that facial feature visibility, combined with eye gaze, appear to work much like a pointing gesture in

some situations. Thus, we have strong support for Hypothesis 4 and Hypothesis 5 in some conditions. We also have strong support for one other result. In some conditions facial feature visibility is effective at increasing the attention paid to artifacts while in other conditions we found no evidence that facial feature visibility has an impact on artifact attention.

Hypothesis 3: Researchers will have a better understanding about artifacts, how they are used, and the information they contain when gesture is used to refer to those artifacts during a presentation.

Our results provide little evidence to support Hypothesis 3. Our analysis provides little evidence that gesture visibility has an impact on questionnaire scores except in a small number of very specific situations. This is a surprising result. Despite strong results that show gesture visibility has a significant impact on artifact attention, there is little evidence that gesture visibility has a significant impact on the questions that involve the artifacts that are used in the presentation. Although we do need to consider the fact that gesture visibility may not have an impact on understanding in scientific presentations, we also need to consider the possibility that our questionnaire is not an effective measure of understanding (see Section 11.4.4 for a discussion of threats to construct validity in this study).

Hypothesis 6: Researchers will have a poorer understanding about artifacts, how they are used, and the information they contain when facial expression is visible as part of the presentation.

Our results show that there is a moderately significant impact of facial feature visibility on questionnaire scores ($0.05 > p > 0.1$), but only on specific questions. This impact is the opposite of what was predicted in our hypothesis. Our results show that on some questions, facial feature visibility resulted in increased scores. Given that we also found that facial feature visibility sometimes increases artifact attention, this finding is perhaps not as puzzling as it seems.

11.4 Threats to Validity

In any experimental study an important consideration is addressing the possible threats to validity. Validity, in this context, refers to the approximate truth of propositions,

inferences, or conclusions. In this section we discuss the various threats to validity that are relevant to this study.

11.4.1 Threats to Conclusion Validity

Conclusion validity is concerned with the degree to which conclusions about relationships in our data are reasonable. One of the key threats to conclusion validity in our study is the small sample size in each of the cells in our multi-factor analyses. Although the study originally targeted 12 subjects per cell, due to the problems encountered with the eye tracking system, it was necessary to reject 13 subjects during the study (see Section 11.4.4). Thus, cell sizes are smaller than desired, with 8, 9, or 10 participants per cell. Small cell sizes can result in low power in the resulting statistics.

In order to mitigate this threat, we performed non-parametric Kruskal-Wallis tests on the ranks [KW52] to confirm that the significance that we encounter with our parametric analysis (ANOVA) is consistent with the non-parametric analysis. The ANOVA results reported in Section 10.1.2 are reflected in our non-parametric Kruskal-Wallis analysis, with the Kruskal-Wallis mean rank test reporting statistically significant differences in mean ranks across our conditions. The fact that our non-parametric tests do not indicate that our parametric ANOVA results are suspect mitigate the chances that our small sample size has a negative impact on our analysis.

11.4.2 Threats to Internal Validity

Internal validity is concerned with the factors that might impact the causal relationship between the treatment and the outcomes of the study. We consider a number of threats to internal validity that are relevant to this study below.

Personal biases: Global warming is a topic about which many people have strong opinions. Such biases could impact the way individuals answer our questionnaire. This threat is mitigated by the fact that the questions in our questionnaire are carefully worded such that they do not ask for responses that require personal opinion. That is, we ask participants about the content of the presentation and in particular, what the speaker states as important, not what the participants feel is important.

Mortality: Our study had a number of participants who had to be discarded from our analysis because of problems with the Tobii eye tracking system (see Section 8.5.1 and

Appendix 15.1.2). Whenever mortality occurs in a study, the reason for this mortality needs to be considered in terms of whether or not it impacts the causal relationship between the treatment and the outcomes of the study. The threat in this study is that the lack of visual cues may lead to boredom and lack of attention, which in turn may lead to failure in our post-study calibration phase. Although we cannot rule this out as a threat to validity, our analysis suggests that participants are rejected from the study because of an error offset in the tracking rather than their lack of attention to our post-study calibration due to boredom. That is, participants who were rejected consistently attended to an area above the calibration area to which they were asked to attend (revealing the error offset), indicating that their attention was focused but that the eye tracker was reporting inaccurate data. It should be noted that the experimental condition in which there were neither gesture or facial features visible (the NGNH condition) did have the largest number of participants that were rejected (4), but all experimental conditions had at least one participant who was rejected (1, 2, and 4 in the YGYH, YGNH, and NGYH conditions respectively).

Testing effects: The subjects in our study knew they were being tested and that eye tracking was being used to measure their gaze. Such knowledge can change subject behaviour. In particular, subjects might "pay more attention" than they normally would if they were "naturally" watching a presentation. The possible effect of a specific individual biasing the results through attention is partially mitigated by experimental design. Subjects are randomly assigned to conditions, and therefore individual differences should be balanced across conditions. In addition, the impact of testing effects is also mitigated by the fact that although subjects know their gaze is being tracked, they are not made aware of the topic of interest (attention to facial features and artifacts). Thus any biases introduced by testing should not bias either artifact or facial feature attention more than any attention paid to other features in the video.

11.4.3 Threats to External Validity

External validity is concerned with how the results from a study generalize to contexts outside the study, including other populations, other measures, and other situations. This study is targeted at scientific researchers at academic institutions, and therefore our sample of senior research students, faculty, research associates, and post-doctoral

researchers is a reasonable sample of this population. Although spanning many ethnicities, our sample is specific to North American academic institutions and may not apply to academic institutions in other countries. In particular, how gesture is used is known to vary across ethnic groups [Ken04, p. 66], and therefore our study of gesture may not apply to academic research in other countries.

Another threat to external validity is our use of a one-way communication to study a two-way collaboration process. Although limiting participants from interacting with the presenter limits the generalizability of this study to considering more interactive collaboration tasks, this limitation is imposed in order to control for the complexities of interactive collaboration tasks. Using a study format in which study participants are non-interactive observers of a face-to-face dialogue is widely used by social psychologists to study face-to-face interaction, and in particular interaction that involves gesture [BC06]. In addition, although our study considers one-way communication, as we see in Chapter 6 and Chapter 7, research presentations that involve primarily one-way communication are common in research meetings, and are therefore an appropriate modality to study. With that said, it is necessary to be careful not to generalize these results to highly interactive, two-way communications without further research (see Section 15.1.1 for a more detailed discussion of this issue).

Another possible threat to external validity is whether or not the video chosen is representative of the types of communication that occur during scientific collaborations. The video was chosen because it presents a complex scientific research topic and makes extensive use of artifact interaction during the presentation. It is therefore representative of artifact-centric scientific communication in these respects. The video is tongue in cheek (the presenter uses humour and props to make his point), so in that respect it is not typical of a scientific presentation. A humorous video was chosen in order to maintain interest in the topic, but it should be noted that this humour may be a confounding factor There is no evidence that demonstrates whether the humour is effective at maintaining attention and at the same time it is unclear whether or not the humour adds or distracts from the scientific point of the presentation.

11.4.4 Threats to Construct Validity

Construct validity is concerned with the accuracy of the mapping from construct or concept (e.g. understanding) to operationalized measurements of that concept (e.g. questionnaire scores). That is, do our measures effectively capture the constructs of concern in our study? Our study has two main constructs that we measure: attention and understanding.

11.4.4.1 Gaze Fixation as a Measure of Attention

Our operationalized measure of attention is gaze fixation. The cognitive psychology literature has shown that spatial attention and gaze often move in tandem [HK03]. Although it is possible for attention and gaze to be spatially disjoint (covert attention), research shows that when individuals are directed to attend to a specific object, gaze fixates on the object on average 250 ms after the object is disambiguated through speech [TSE+95]. Given that our study considers just such a situation (artifacts are disambiguated through either speech or gesture), gaze fixation is an appropriate measure of attention. Threats to the validity of this measure stem from two directions.

First, we must determine whether the Tobii eye tracker is effective at measuring what the eye is fixated upon. The accuracy of the Tobii eye tracker (is it actually measuring where the eye is fixated on the screen?) was an issue in this study (see Section 8.5.1 and Appendix 15.1.2). We use pre-study and post-study calibration phases that allow us to reject subjects for whom eye tracking is an issue. Although our pre-calibration phase eliminates all participants for which the Tobii eye tracker is not effective, we found that for some participants the eye tracker started to report fixations with an erroneous fixation offset part way through the study. We identified these cases in the study through the post-study tracker calibration phase at the end of the movie for each participant. Those participants for whom the eye tracking offset occurred were discarded from the study. It is worth noting that although most participants who were rejected using the post-study calibration clearly exceeded our rejection criteria (see Appendix 15.1.2), there were two participants whose post-study calibration was borderline in terms of our rejection criteria. One of these participants narrowly exceeded our rejection criteria (and was therefore rejected) and one participant narrowly exceeded the criteria (and was therefore used in the study).

Another potential threat to the validity of our measure of attention is that the Tobii eye tracker reports fixations with gaps between fixations. The length of time between fixations varies based on participant. Although gaze fixations are typically spatially coherent, during the times between fixations we have no data on where the participant's gaze is fixated (see Section 8.7.1 for more details). Although we have no evidence that gaze fixations are spatially disjoint (that participants are attending to a different part of the screen), this eventuality cannot be eliminated.

11.4.4.2 Questionnaire Score as a Measure of Understanding

Understanding is particularly difficult to measure. Our operationalized measure for understanding is questionnaire score. Our analysis (see Section 11.2 and Section 11.3) provides little to no support for our hypothesis that gesture visibility has an impact on understanding. Pre-study testing of our questionnaire identified that some of the questions may be too easy (scores were very high), resulting in the modification of some of the questions used in the study. Despite these changes, our questionnaire did not capture differences across our conditions. Although this may suggest that our experimental conditions do not have an impact on questionnaire score, the significant impact of our experimental conditions on artifact attention suggests that our operationalization of the concept of understanding may be suspect.

We hypothesize that this result may be due in part to questionnaire design. The scores on individual questions were either very high or very low. The questions that were difficult were consistently difficult. For example, 40 out of 50 participants got 1 out of 2 on Question 2 and 47 out of 50 got either 1 or 2 out of 5 on question 5a. This lack of spread in the scores raises questions as to whether the questionnaire is effectively measuring the effect of our experimental intervention. It may be the case that Question 2 and Question 5a, as posed in the questionnaire, were simply too difficult to get good scores given the presentation material. For example, Question 5a asks participants to identify the five "global disasters" listed in one of the diagrams used in the presentation. Our analysis indicates that obtaining correct answers to this question may have been confounded by the fact that the presenter provided a range of "global disasters" throughout the talk but were not one of the five "global disasters" that were specifically asked for in Question 5a. Thus, participants were faced with the task of remembering five

out of approximately twenty alternatives (see Section 10.2.2 for a more detailed discussion).

The questions that were easy were also quite easy. When considering individual participant scores on individual questions (325 data points), 82% of the individual question scores were given full marks (2 out of 2). Even in the no gesture visibility and no facial feature visibility condition (NGNH), where no human communication channels are visible, 82% of the questions were answered 100% correctly. We are left with a situation in which it is unlikely that our experimental interventions would have a dramatic impact on questionnaire scores, simply because scores are very high even when there are no visible facial feature or gesture cues.

Although our results do not allow us to state that gesture visibility has an impact on questionnaire scores, given the strong evidence that gesture draws attention to artifacts we hypothesize that the lack of increase in questionnaire scores (and in turn understanding) may be a result of the fact that the questions posed were not able to capture the impact of our experimental conditions. Of course, it is also important to point out that these results may indeed suggest that gesture does not have an impact on understanding. More research is required in this area.

11.5 Conclusions

We summarize our results from this study as follows:

- We have strong evidence that gesture visibility increases the attention paid to artifacts. This is a strong and consistent result across almost all conditions.
- We have strong evidence that facial feature visibility sometimes decreases the attention paid to artifacts, sometimes has no impact on artifact attention, and sometimes increases the attention paid to artifacts. This is also a strong result, but shows that facial feature visibility is more complex in its effects on artifact attention.
- We have little to no evidence that gesture visibility has an impact on questionnaire scores. We show no impact on questionnaire score, but have some reasons to question whether our measures on questionnaire score tested understanding effectively. More research is required.

- We have only moderate evidence that facial feature visibility has an impact on questionnaire scores. Interestingly, the impact of facial feature visibility increases, not decreases, the scores. Statistically, this is not a strong result but combined with our results on the effect of facial feature visibility on artifact attention, this is an interesting result. This is particularly true in the context of the lack of results that we see on the impacts of gesture visibility on questionnaire score. Again, more research is required in this area.

We consider the results of this study in the context of the overall framework of this dissertation in Chapter 12.

Part IV – Summary

12 Design Guidelines

One of the primary objectives of this research is to take the knowledge generated from our studies and to "*Develop a set of design guidelines for the development of effective collaboration tools for scientific researchers*." (Objective 4:). In particular, we attempt to answer the basic research question *"What human communication channels need to be supported for artifact-centric collaboration?"*

Our research demonstrates several high-level trends. First, collaboration technologies are an important tool in the support of modern science. Our study of SMS usage at the IRMACS Centre shows that researchers make extensive use of such collaboration technologies when they are made available (Chapter 6). Second, researchers make extensive use of digital artifacts during research meetings and that they use gesture extensively to refer to those artifacts (Chapter 7). Third, both artifacts and facial features are attended to extensively during scientific collaboration and that both gesture and facial features are successful at drawing attention to artifacts (Chapter 8 through 11). Fourth, modern distributed collaboration technologies do not effectively communicate gesture to remote participants (Chapter 7). It is these findings upon which we base our design guidelines. We divide our design guidelines into two types: guidelines for collaboration tool builders and guidelines for collaboration infrastructure designers.

12.1 Guidelines for Tool Builders

Our guidelines for tool builders are targeted at software developers or collaboration support personnel who are responsible for creating and/or deploying advanced collaboration software in the support of distributed, artifact-centric, scientific collaboration.

12.1.1 Supporting Shared Access to Digital Artifacts

Tools that support distributed, artifact-centric, scientific collaboration should support shared access to digital artifacts. We begin by extending one of the collocated, tabletop collaboration guidelines as specified by Scott *et al.* [SGM+03] to the distributed artifact-centric collaboration domain. Our studies shows that researchers often have rich interactions with digital artifacts, both collocated and distributed, and that it is necessary to provide the ability for any participant to interact with and control the digital

workspace. In order to support shared access to digital artifacts, collaboration tools should support the ability to interact directly with these artifacts. These interactions should be as reciprocal as possible for both collocated and distributed participants. That is, all participants should be able to interact with the digital artifacts in the same way.

12.1.2 Support Natural Artifact Interaction

Tools that support distributed artifact-centric, scientific collaboration should support natural artifact interaction. Technology used to support distributed, artifact-centric scientific collaboration needs to capture and communicate natural user interactions with digital artifacts. In this instance, we define a natural interaction as one that does not have to be adapted to technology to communicate effectively with remote collaborators. Our results support related research that shows that when people interact with a screen that is displaying digital artifacts, their interactions are natural and fluid [SGM03]. These natural interactions are facilitated by the ability of the display surface (e.g. Smartboard or tabletop interaction devices) to support direct manipulation of the artifacts.

In the domain of scientific collaboration, our analysis shows that gestures are used extensively to both refer to and manipulate artifacts. Many of those interactions are performed with natural physical gestures (made with the hand rather than interacting with the Smartboard or the mouse). Our research also shows that such gestures, when they are communicated, are highly attended to by researchers while watching research presentations. In addition, our studies suggest that the physicality of the interaction environment (co-location of the display and the interaction device) may also encourage such interaction. A high percentage of physical pointing gestures are performed when collocated with the display. Multi-person interaction also tends to occur more often when collocated with the display.

Our results suggest that this natural tendency to point physically is a compelling one. In our studies, participants often forgot that they needed to use the technology (the mouse or the Smartboard) to communicate gesture effectively. Note that they typically did not stop gesturing, but instead transitioned from using technology based gestures (which are transmitted to remote participants) to physical gestures (which are not transmitted to remote participants). In order to effectively communicate such artifact gestures, tool builders have two choices. We can develop more effective ways of using existing

interaction technologies so that users can more easily adapt to their use, or we can develop new gesture communication technologies that capture natural gesture and communicate it effectively.

Our research suggests that we should do both. Tool builders should try to capture and communicate natural artifact interactions as effectively as possible. This is the "holy grail" of distributed, artifact-centric collaboration. The use of high definition video and image processing are promising technologies in this regard. At the same time, our research shows that users will adapt to using technology to communicate about artifacts and that this adaptation will occur rapidly when the artifact interaction is critical to the collaboration task. Thus, tool builders should take this adaptation into consideration when designing distributed, artifact-centric collaboration tools. Tool builders should minimize the amount of adaptation required by users through facilitating natural artifact interactions and merging personal interaction and task interaction spaces.

12.1.3 Supporting Interpersonal Interaction

Tools that support distributed, scientific collaboration should support interaction with other people. Building on the design guidelines from Scott *et al.* on collocated tabletop interaction [SGM+03], our results show that such interpersonal interaction is important in both collocated and distributed artifact-centric collaboration. Our results suggest that interpersonal interaction in front of a vertical display surface that supports direct touch interaction (as opposed to a tabletop environment) is an important modality for collocated artifact-centric, scientific collaboration. Our results also show that this modality should be extended to distributed participants as well. Such interpersonal interaction is critical in providing work space awareness and in the ability to locate what others are doing, align work with others, and control and monitor access to artifacts.

One of the surprising outcomes of our gesture decoding study is that facial feature visibility does not always have a negative impact on artifact attention, and in some cases facial feature visibility increases the attention paid to artifacts. Although more research needs to be performed in this area, our findings suggest that providing facial feature visibility to distributed participants can be useful in supporting artifact-centric collaboration. Note that our research only suggests that facial feature visibility is helpful in an integrated personal and task space where proximity and gaze direction are

consistent with the task space. It is important to point out that this research does not provide any information (positive or negative) as to the value of the picture-in-picture (PIP) style of facial feature visibility. PIP facial feature visibility, where a visual stream of a participant exists in one corner of the screen (not collocated with task space), is common in most video conferencing technologies. Our explorations focus on the integration of personal and task space, and therefore our results that suggest facial feature visibility increases artifact attention only apply when facial features are integrated with task space.

It is also worth noting that the integration of facial features and gestures into a single integrated task space is facilitated by the affordances of a vertical display (as compared to a tabletop display). That is, we stand in front of a wall mounted screen and our gesture and facial expressions are collocated with the task environment. An example of this can be seen in our gesture decoding study where we were able to provide a facial feature visibility condition in which facial features were collocated with the work space. Thus, the ability to create a work space with integrated facial features and gestures is feasible in a vertical screen interaction environment. Such a system would be difficult or impossible in a tabletop collaboration environment. Given that a researcher's work often revolves around a whiteboard style environment, the affordances of wall mounted displays suggest a promising avenue for supporting distributed, artifact-centric scientific collaboration.

12.2 Guidelines for Infrastructure Builders

Our guidelines for infrastructure builders come largely from our analysis of the operation of the IRMACS SMS infrastructure from 2005 to 2010. Although these guidelines are based on our SMS analysis, the reader should be aware that they are also influenced by the author's experiences in developing, deploying, operating, and supporting this infrastructure during this time period. Given that longitudinal studies of operational scientific collaboratories are almost non-existent (especially spanning a five year period), it is hoped that the reader sees the value in using these experiences to supplement the development of these guidelines. As with any guidelines, infrastructure builders should consider each guideline carefully to determine if it applies to their situation.

12.2.1 Distance Matters

Collaboration infrastructure that supports researchers should be located as close as possible to the user community. There has been extensive research that shows that distance matters when supporting collaboration [OO00]. Our observations reflect this as well. Our analysis of SMS usage at the IRMACS Centre suggests that the amount of collaboration usage is at least partially related to the size of the research community at the centre. The IRMACS Centre puts approximately 200 researchers in close proximity with five advanced SMS rooms. Our experience with operating this infrastructure over the last five years suggests that this collocation lowers the barrier to researchers adopting and using distributed collaboration technologies. On a daily basis, researchers hold meetings in these rooms, and it is as easy for such a meeting to be a distributed meeting with colleagues at another institution as it is with their local research group. Users of IRMACS SMS rooms clearly state the importance of proximity to SMS environments.

12.2.2 Flexibility and Extensibility

Collaboration infrastructure that supports scientific collaboration should be flexible and extensible. Our analysis of the SMS collaboration infrastructure at the IRMACS Centre shows that SMS rooms are used for a wide range of purposes. No one hardware or software technology will meet the research needs of a wide research community. Thus, an SMS room that supports a general research community should be designed to be as flexible and extensible as possible. It should support a wide range of software technologies and the hardware should be reconfigurable such that it can be easily configured to meet a specific researcher's collaboration needs.

12.2.3 Ease of Use

Collaboration infrastructure that supports a research community should be easy to use. The large number of SMS sessions that are held within the IRMACS facilities are indicative of how important distributed collaboration is to IRMACS researchers. In order to facilitate repeated use of advanced SMS technologies, it is important that the infrastructure be easy to use. Note that flexibility (as discussed above) and ease of use are often in conflict, but our experience suggests that this need not be the case. In supporting researchers at IRMACS, there are a relatively small number of common collaboration use

cases. With a small number of use cases, it is possible to design the technical infrastructure to support a limited set of collaboration scenarios such that they are straightforward to use. That is, an SMS room can be configured to support a given scientific collaboration scenario at the touch of a button (or two). SMS infrastructure designers and developers should have a good understanding of the research scenarios that are important to their research community before designing and deploying an SMS environment.

12.2.4 Supporting Fluid Transitions between Activities

Distributed, scientific collaboration infrastructure should support a fluid transition between research activities. We again return to the tabletop design guidelines presented by Scott *et al.* [SGM+03], extending them to include distributed, artifact-centric collaboration. As pointed out by Scott, tabletop interaction should support fluid changes between activities. Our analysis of research meetings (our ethnography) shows that such meetings rapidly change phases and that different phases often incorporate different activities or tasks. We have observed meetings in which activities change as often as once per minute (on average). In addition, our analyses indicate that not all phases include artifact interaction. Thus, not only is it necessary to support different artifact interaction activities effectively, but it is also necessary to support non-artifact interaction.

It is worth noting that this design guideline could just as easily be listed under guidelines for tool builders. Transition between activities is an issue for tool builders, as the software tools need to support changes in the type of artifact interaction (referring to artifacts versus creating and modifying artifacts) as well as changes between activities that require artifact interaction and those that do not. It is also an issue for infrastructure builders as the design of the technology, such as the relative location of displays and touch interaction technologies to tables and chairs, needs to be taken into consideration.

Note that this is much more problematic when vertical displays are being used for collaboration than when a tabletop is being used. That is, in a vertical display environment, a discussion phase that does not involve artifact interaction typically results in everyone sitting down around a meeting room table, while activities that involve artifact interaction often result in one or more people standing up at the display. Thus, when supporting artifact-centric collaboration in a touch sensitive, wall mounted display

environment, it is important to consider the affordances of transitioning between artifact interaction activities where users are standing at the display and non-artifact interaction activities where users are sitting at a table. Note that this is primarily an issue with the physical environment in a room as opposed to a distributed collaboration issue. The affordances of fluid transition between activities from a distributed collaboration perspective are primarily a collaboration tool builder issue and in many ways are taken into consideration in the design guidelines for tool building (our ability to access artifacts, interact naturally with artifacts, and perform inter-personal interaction).

13 Conclusions

In Chapter 1 we outline the importance of distributed, artifact-centric collaboration tools to the scientific research community. This importance is a result of the confluence of several factors: the increasing importance of computational science as a method of research; the data deluge that scientists are currently experiencing due to access to large experiments, high-resolution instruments, and large scale computational simulations; and the globalization of research teams as researchers attempt to find the required expertise to solve today's complex scientific problems. We return to Hamming's 1962 quote *"Computing is about insight, not numbers"*. The goal of this research is to facilitate researchers achieving insight by helping to bring the right researchers together, at the right time, with the right data. This is the domain of distributed, artifact-centric scientific collaboration.

The high-level goal of this research is to gain a better understanding of the impact that distance has on distributed, artifact-centric, scientific collaboration. In order to achieve this goal, a set of four objectives were laid out:

Objective 1: Develop a broad understanding of how scientific researchers collaborate.

Objective 2: Develop a deep understanding of how scientific researchers interact with digital artifacts when they collaborate.

Objective 3: Evaluate advanced collaboration modalities and technologies for scientific collaboration.

Objective 4: Develop a set of design guidelines for the development of effective collaboration tools for scientific researchers.

We explore how each of these objectives was met below.

13.1 Addressing the Objectives

13.1.1 Broad understanding of how scientists collaborate

> *Objective 1: Develop a broad understanding of how scientific researchers collaborate.*

The focus of our research is at the intersection of several complex research areas, including computer supported collaborative work, communication, social psychology,

and cognitive psychology. Thus, a broad understanding of the domains of interest is necessary. We fulfill this objective by using a multi-dimensional research approach where we triangulate on research questions by considering more than one level or technique. We use both qualitative and quantitative research methods and measures throughout our study, and consider the domain of distributed, artifact-centric collaboration across the following dimensions:

- We study distributed scientific research at both the macro (broadly over a large research centre – Chapter 6) and micro (detailed within in a single research group – Chapter 7) levels;
- We study both collocated (Chapter 5 and 7) and distributed research groups (Chapter 5 through 10);
- We study both the encoding (how information is sent - Chapter 5 and 7) and decoding (how information is received – Chapter 8 through 10) processes researchers use to communicate about complex scientific topics; and
- We study state-of-the-art collaboration technologies in both research prototype (Chapter 5) and production environments (Chapter 6 and 7).

Although current literature provides a number of theories, models, and frameworks that attempt to capture the complexity of this problem domain, a sufficiently comprehensive and cohesive framework that brings these fields together has been elusive. In order to capture this complexity and further our analysis, the CoGScience Framework was created. The CoGScience Framework (Chapter 4) captures the broad scope of this problem domain, while at the same time providing us with a tool to perform detailed analyses of specific collaboration scenarios. We utilize this framework in all aspects of our analyses throughout this dissertation.

Last, but certainly not least, our longitudinal study of the use of Scientific Media Spaces (SMS) at the IRMACS Centre from 2005 – 2010 provides us with a high-level, broad view of how a scientific community makes use of advanced collaboration technologies (Chapter 6). Our research indicates that the size of the research community, the level of research activity, and the availability of SMS technologies all have an impact on the frequency of use collaboration technology.

13.1.2 Deep understanding of how researchers interact with data

Objective 2: Develop a deep understanding of how scientific researchers interact with digital artifacts when they collaborate.

Reaching the objective of having a deep understanding of artifact-centric collaboration is challenging. We achieve this objective through three key studies: CoTable (Chapter 5), our longitudinal ethnographic study (Chapter 7), and our laboratory study of gesture decoding (Chapter 8 through 11).

The complexities of providing a functional, artifact-centric, distributed, collaboration environment became apparent during our study of the CoTable system (Chapter 5). Our study of the CoTable prototype suggested that **gestural** interaction with digital artifacts is a critical communication **channel**. More importantly, it suggested that the richness of application-based artifact gestural communication **visual streams** (traditional application based tele-pointers) do not provide enough human communication **channels** (**gesture**, **body language**, **workspace awareness**). In particular, our observations suggest that users prefer a low **fidelity** visual **stream** that provides a richer set of communication **channels** over a high **fidelity** visual **stream** that provides few communication **channels**. Although CoTable was a simple case study analysis, the somewhat surprising results led to the main objectives considered in this research.

The longitudinal ethnographic study presented in Chapter 7 answers many of the questions raised by our CoTable exploration. It provides us with a deep understanding of how scientific researchers work with digital artifacts in both collocated and distributed environments. In particular, our findings demonstrate that the frequency of artifact gestural interactions in scientific collaboration is as high as it is in other artifact-centric application domains [BOO95, TL99, Tan89]. Our findings indicate that gesture is used extensively in some phases of a research meeting (but not in all phases). In addition, our results show that artifact gestural interaction is used in both **loosely coupled description** phases as well as in **tightly coupled discussion** or **exploration** phases. In particular, artifact gesture interaction appears to be used frequently when there is **competitiveness**, **conflict**, **urgency**, and **excitement** involved in communication about the artifact. Our findings also indicate that artifact gesture interaction may be facilitated by touch interaction devices like the Smartboard, as such devices encourage having people in close

proximity to each other (encouraging inter-personal interaction) and to the artifacts themselves (encouraging artifact interaction).

Our ethnography provides us with a deep understanding of how gesture is used to communicate information from an encoding perspective. That is, it tells us how researchers use gesture to communicate information to others when they interact with digital artifacts. Although the ethnography does not tell us anything about how researchers attend to such artifact gestures, it does lead to a set of questions regarding attention and artifact interaction. These questions, in turn, became the hypotheses we explored in our laboratory based gesture study (Chapter 8 through 10).

The laboratory gesture study provides us with a deep understanding of the decoding process of artifact-centric scientific communication. In particular, it considers the impact of gesture and facial feature visibility on attention and understanding during a scientific presentation.

The study provides strong evidence that gesture visibility increases the attention paid to digital artifacts, and this evidence is consistent across almost all conditions. Although there is strong evidence that facial feature visibility sometimes decreases the attention paid to artifacts, this is not true in all conditions. There is also strong evidence that in some conditions, facial feature visibility results in an increase in the attention paid to artifacts. Of particular interest is the fact that facial feature visibility appears to work much like a pointing gesture in some situations, drawing attention to artifacts through proximity and eye gaze direction.

Our analysis shows that gesture visibility has little impact on understanding (as measured by our questionnaire) except in a small number of very specific situations. Although gesture visibility appears to have no effect on understanding, there is some evidence that our questionnaire scores were not a suitable measure of understanding. More research needs to be performed in this area. Our results show that on some questions, facial feature visibility has a moderately significant impact on increasing study questionnaire scores, rather than decreasing scores. Again, this impact occurs only on some questions. Although this result is not what was expected, it is consistent with our findings on artifact attention. That is, facial feature visibility sometimes increases and sometimes has no effect on questionnaire scores.

The results on facial feature visibility (both in attention and questionnaire scores) are interesting to consider. Our original hypotheses suggested that facial feature visibility would distract participants from attending to artifacts and reduce their understanding about those artifacts. Although we have support for this hypothesis in some conditions, we have support for the opposite effect as well. This implies that it may be possible to integrate gestural visibility and facial feature visibility into a single artifact-centric collaboration workspace without having a dramatic negative effect on the attention paid to artifacts in many collaboration scenarios.

13.1.3 Evaluate advance collaboration modalities and technologies

Objective 3: Evaluate advanced collaboration modalities and technologies for scientific collaboration.

The main intention of the CoTable prototype collaboration system (Chapter 5) was to explore and experiment with advanced technologies in a rich, face-to-face collaboration environment. The tabletop environment was chosen because it was a natural, multi-user environment that is rich in subtle communication channels. Although providing a relatively complex and rich collaboration environment, our experiences with CoTable revealed that a rich set of communication channels was necessary, but not sufficient, to providing a compelling collaboration environment. In particular, the apparent willingness of users to utilize a rich but low fidelity visual stream in preference over a less rich but higher fidelity visual stream suggested that capturing rich, multi-modal communication in an effective way is key to providing an effective, artifact-centric collaboration environment.

Our study of the SMS rooms provided by the IRMACS Centre (Chapter 6) demonstrates that advanced collaboration technologies are useful tools to the scientific research community. With over 486 remote meetings (27% of the total meetings) linking remote sites to the IRMACS Centre in 2009, the impact of remote technologies on the broad scientific research community is clear. Our ethnographic study reveals a similarly strong impact of the technology on a single research group. Collaboration technologies, and in the case of our ethnography, the artifact-centric capabilities of the technology, are considered important, and often critical, components of the groups work process. Participants in our focus group state that:

> *"… now we can not live without these things [Smartboards]"*

> *"The fact that [the remote collaborator] was at home made no difference, we were all around the Smartboard, doing the same thing."*

At the same time, it is clear that the technologies have a long way to go before they provide a complete solution for distributed, artifact-centric, scientific research. Even the advanced collaboration tools utilized with CoTable, those deployed in the IRMACS Centre, and those studied in our ethnography are unable to capture the subtle interactions that are necessary to effectively communicate artifact interactions. In particular, our ethnography demonstrates that despite our participants thinking of the artifact centric collaboration as "*brilliant*" in terms of its success, our findings indicate that a very large percentage of artifact gestures are not effectively communicated to remote participants. We hypothesize that touch sensitive screens, and the physical interactions that accompany them, facilitates and encourages physical interaction both among users and with digital artifacts. Thus, providing advanced interaction technologies encourages the use of a rich, artifact-centric collaboration space, which is exactly the space that our collaboration tools fail to support effectively.

In addition, it is important to note that both the SMS analysis of the IRMACS Centre and our ethnography indicate that artifact-centric collaboration is not the only barrier to distributed, scientific collaboration. Even the simplest of technologies (e.g. the telephone) can break down and result in an entire meeting grinding to a halt because a key participant is prevented from attending. Tools are still difficult to use and often unreliable, making the barrier to scientific collaboration relatively high. Although users do adapt to the use of collaboration technologies, the level of adaptation required is still far too high. This is especially true of advanced technologies such as sharing applications with remote collaborators and interacting with touch screen technologies.

> *"Because our interactions always have as a focus either a document, looking at, commenting on, creating, and that document is on the smart board, the whole texture of the meeting is incredibly sensitive to how well the smart board technology and the document manipulation works. Any glitches […] send things off the rails so quickly. We just lose momentum, which is a disaster."* Focus group participant

13.1.4 Develop a set of design guidelines

Objective 4: Develop a set of design guidelines for the development of effective collaboration tools for scientific researchers.

By considering the range of experiences and analyses carried out as part of this dissertation, a set of design guidelines for the creation, deployment, and operation of advanced, artifact-centric scientific collaboration environments has emerged (Chapter 12). These guidelines are divided into guidelines for tool builders and guidelines for infrastructure builders. The guidelines suggest that it is important that artifact-centric collaboration tool builders support shared access to digital artifacts, support natural artifact interaction, and support interpersonal interaction. The guidelines suggest that infrastructure builders consider proximity of the infrastructure to the research community, the flexibility and extensibility of the infrastructure to meet a wider variety of researcher collaboration needs, the ease of use of the infrastructure, and the way the infrastructure supports fluid transitions between research activities.

13.2 Contributions

A number of research contributions have emerged from the research presented in this dissertation.

13.2.1 Empirical CSCW Contributions

This dissertation presents the results from four research studies, a case study analysis of an advanced artifact-centric tabletop collaboration prototype, a case study analysis of the broad SMS usage in a large research centre, an in-depth longitudinal ethnography of the encoding process of scientific collaboration, and an in-depth laboratory study of the decoding process of scientific communication. In particular, our longitudinal studies of the broad research community and an individual working research group in a naturalistic, yet high technology, collaboration environment provide us with a unique perspective on how scientists collaborate. All of these studies provide new empirical evidence that helps to address the questions posed by this research.

13.2.2 Empirical Social Psychology Contributions

Our laboratory study of the decoding process in artifact-centric scientific collaboration also presents new quantitative results in social psychology. The decoding process of

gestural interaction is difficult to study, with most gesture studies using indirect measures of gesture movement combined with dialogue to measure attention and communicative meaning. Our results contribute new empirical evidence as to the importance of both gesture and facial feature visibility to the attention paid to referent artifacts during conversation. To our knowledge, our laboratory gesture study is the first study to utilize eye tracking to consider the impact of gesture visibility on attention in artifact or object-centric communication.

13.2.3 Gesture Coding Methodology

The technique developed for coding gesture, although based on the foundations of other gesture coding schemes, provides a new way of considering gestural interaction (see Section 7.1.2 and Section 7.1.3). In particular, the approach that we take to combine low-level communicative actions such as utterances and gestures into higher-level gestural communication events allowed us to differentiate between three different types of gesture events. This also allowed us to consistently consider these three types of high-level gesture events across both our ethnography and our laboratory study.

13.2.4 CoGScience Framework

The synthesis of our studies and experiences also resulted in the creation of the CoGScience Framework. The framework bridges the gap between the human communication needs of artifact-centric scientific collaboration and the technological aspects of how those communication needs can be delivered. The framework utilizes a broad basis of research at its foundation, including research from the social sciences (communication, social psychology, and cognitive psychology) and computer science (CSCW and HCI). The CoGScience Framework can be used as a tool for researchers, designers, or software implementers to rigorously explore artifact-centric collaboration. It has been used throughout this dissertation in exactly these roles.

13.2.5 Design Guidelines

The synthesis of the empirical results discussed above, combined with our experiences in developing, deploying, and operating the IRMACS SMS infrastructure have enabled us to develop a set of design guidelines. These guidelines are targeted at both the

developer of advanced artifact-centric collaboration tools as well as those that might build, deploy, and operate advanced collaboration facilities.

13.3 Future Work

Although the research presented in this dissertation presents new understanding about the impacts of distance on distributed, artifact-centric, scientific collaboration, there is still significant research that needs to be performed.

13.3.1 Study of Wall Mounted Touch Screen Distributed Collaboration

The domain of distributed, artifact-centric, scientific collaboration is complex, and our empirical studies are only a start to the exploration of this area. There are still many unanswered questions. In particular, our ethnography resulted in the creation of a number of hypotheses regarding artifact-centric collaboration. Some of these hypotheses were addressed in the gesture study, but some remain untested. Our results suggest that a more detailed exploration of the impact of physical co-location of people and interaction technology on artifact interaction would be fruitful. In particular, our analysis suggests that touch sensitive wall mounted display surfaces may provide some affordances for integrating personal and task space in distributed, artifact-centric collaboration. Such interaction has been suggested as a significant benefit of tabletop interaction for some time [SGM+03][TPI+10], but the same level of research has not been performed for touch interactive wall mounted displays.

13.3.2 Study of Collaboration in the Computational Sciences

Our study of the use of SMS technologies in the IRMACS Centre, although providing new information about how researchers use advanced collaboration technologies, is only a beginning. The research literature in this area is very sparse, with little existing research into the use of advanced collaboration technologies as part of the scientific research process. In particular, there is a significant gap in the literature as to how collaboration technologies are used and adopted in scientific research communities. Such investigations should include both the evaluation and development of tools that support scientific collaboration as well as the evaluation of how such tools are adopted in the broader computational science community.

13.3.3 Improving Tools for Scientific Collaboration

The design guidelines presented in this research suggest several key directions for creating successful artifact-centric collaboration tools. The only way to validate these design guidelines is to build tools that follow these guidelines and evaluate their success. In many ways, the design guidelines capture the holy grail of artifact-centric collaboration – that is, the ability to create a coherent distributed reference space that effectively integrates personal interaction space with the task space of the digital artifacts. It is worth noting that there is a convergence of technology innovations that may make such a workspace feasible. The availability of commodity high definition video conferencing at a resolution of 1920x1080 pixels implies that an artifact-centric collaboration environment can utilize a single visual stream to capture both a high fidelity task space (laptop resolution of 1024x768 pixels) and a high fidelity personal space. Such a visual stream, combined with today's advanced networks and evolving touch screen user interaction technologies, implies that it may be possible to build such tools effectively in the near future. This is particularly true in the scientific research domain, where such technologies are becoming increasingly common.

13.3.4 Study of the Impact of Gesture on Understanding

Our gesture decoding study also left some questions unanswered. The results on the impact of gesture visibility on understanding were inconclusive and require further research. Although our results on the impacts of facial feature visibility on attention and understanding provide us with important new insights, the impacts are complex and require further exploration. And finally, our gesture decoding study did not consider decoding in the context of interactive communication. Thus, further studies that consider the artifact-gesture decoding process in an interactive collaboration context would complement this research.

13.3.5 Evaluate and Refine the CoGScience Framework

The CoGScience Framework distils research from a broad range of areas into a single framework that encompasses artifact-centric collaboration. Our experiences in applying the framework to the research presented in this dissertation have been very positive. We have found that CoGScience is a useful tool for evaluating existing tools, planning

studies, designing experiments, and analyzing the results from our studies. In order to validate the framework's usefulness as a tool for studying artifact-centric collaboration, it is necessary to apply the framework in a broader context. Applying the framework to the evaluation of other bodies of artifact-centric collaboration research, to the analysis of a broader set of artifact-centric collaboration tasks, to the evaluation of specific artifact-centric collaboration solutions, and to the development of new artifact-centric collaboration tools are all important in validating the usefulness of the framework. Such evaluations will undoubtedly lead to further refinements as the framework evolves to meet researcher needs in artifact-centric collaboration.

13.4 Final Summary

Scientific research is fundamentally collaborative in nature. Two recent trends in this area have transformed the way researchers work together. First, modern scientific instruments and computational simulations are producing data at an unprecedented rate. Second, distributed research groups are becoming common place, as researchers make use of digital technologies to bring together research expertise from institutions across the country or around the world. Today, scientific insight is about bringing the right people together, with the right data, at the right time. This is the domain of distributed, artifact-centric collaboration.

There is an emerging need and opportunity for research in this area. Scientific collaboratories are becoming common, and yet their needs are poorly understood. Computational science is producing data at an unprecedented rate, and yet effective artifact-centric collaboration tools are essentially non-existent. Gestural interaction is seeing a resurgence in research interest in the social psychology community, but it is not supported in remote collaboration tools. Touch sensitive devices are becoming ubiquitous (from the phone to wall displays), and yet we have failed to develop compelling collaboration tools that make use of them.

This research addresses these issues in three fundamental ways. First, our multi-dimensional research approach allows us to triangulate on the importance of artifact interaction in collaborative science. We do this through a number of studies, including case studies, an ethnography, and a laboratory experiment. Second, the CoGScience Framework, which emerged from our study of artifact-centric collaboration, provides us

with a conceptual lens with which to consider artifact-centric collaboration. Third, the contributions from this research include a coherent set of empirical results and a set of design guidelines that suggest mechanisms for the creation and deployment of distributed, artifact-centric collaboration tools.

Although each of the empirical results discussed above is an important contribution in its own right, these contributions are only a small piece of a much larger puzzle. The real challenge in supporting distributed, artifact-centric scientific collaboration is in distilling these, and other related research results, into a coherent view of this problem domain. The multi-dimensional approach used in this research uses our empirical contributions as building blocks in the creation of a strong foundation in this area. The CoGScience Framework stands on this foundation, providing a conceptual scaffolding on which to build a detailed understanding of distributed, artifact-centric, scientific collaboration.

14 Bibliography

[Atlas] www.atlas.ch

[AMJ+02] Anderson, A., Mullin, J., Jackson, M., Smallwood, L., Sasse, A., Wilson, G., and Watson, A. (2002) Audio & Video Guidelines for Networked Multimedia Applications: Applying the ETNA Taxonomy, Downloaded December 7, 2009. www-mice.cs.ucl.ac.uk/multimedia/projects/etna/avguidelines.pdf

[ABI+97] Angiolillo, J., Blanchard, H., Israelski, E., and Mane, A. (1997) Technology Constraints of Video-Mediated Communication, in *Video Mediated Communication* (K. Finn, A. Sellen, and S. Wilbur editors), Lawrence Erlbaum and Associates, Mahwah, USA.

[AMM+03] Apperley, M., McLeod, L., Masoodian, M., Paine, L., Phillips, M., Rogers, B., and Thomson, K. (2003) Use of video shadow for small group interaction awareness on a large interactive display surface. In *Proceedings of the Fourth Australasian User interface Conference on User interfaces 2003 - Volume 18* (Adelaide, Australia). R. Biddle and B. Thomas, Eds. ACM International Conference Proceeding Series, vol. 36. Australian Computer Society, Darlinghurst, Australia, 81-90.

[Bae93] Baecker, R. (1993) *Groupware and Computer-Supported Cooperative Work*, Morgan Kaufmann: San Francisco, pp 165-168

[BC79] Bales, R. and Cohen, S. (1979) *SYMLOG: A System for the Multiple Level Observation of Groups*, Collier Macmillan.

[Ban04] Bangerter, A. (2004). Using Pointing and Describing to Achieve Joint Focus of Attention in Dialogue. *Psychological Science*, *15*(6), 415-419

[BC07] Bangerter, A. and Chevalley, E. (2007). Pointing and describing in referential communication: When are pointing gestures used to communicate? *CTIT Proceedings of the Workshop on Multimodal Output Generation (MOG),* (I. Van der Sluis, M. Theune, E. Reiter, and E. Krahmer editors), Aberdeen, Scotland, January 2007.

[Bar68] Barnlund, D. C. (1968) *Interpersonal Communication: Survey and Studies*. Boston: Houghton Mifflin, 1968.

[Bar32] Bartlett, F. (1932). *Remembering: An Experimental and Social Study.* Cambridge: Cambridge University Press.

[BGS08] Bavelas, J., Gerwing, J., Sutton, C., and Prevost, D. (2008) Gesturing on the telephone: Independent effects of dialogue and visibility, *Journal of Memory and Language*, 58 (2), pp 495-520, Elsevier.

[BG07] Bavelas, J. and Gerwing, J., (2007) Conversational Hand Gestures and Facial Displays in Face-to-Face Dialog, *Social Communications* (ed. K. Feidler), pp 283-308, Psychology Press [*Frontiers of Social Psychology* Series]. New York.

[BC06] Bavelas, J. and Chovil, N. (2006) Nonverbal and Verbal Communication: Hand Gestures and Facial Displays as Part of Language, *Handbook of Nonverbal Communication* (eds. V. Manusov and M. Patterson), pp. 97-115, Sage, Thousand Oaks, CA.

[BC00] Bavelas, J. and Chovil, N. (2000) Visible Acts of Meaning: An integrated Message Model of Language in Face-to-Face Dialogue, *Journal of Language and Social Psychology*, 19(2), 163-194.

[BCC+95] Bavelas, J. B., Chovil, N., Coates, L., and Roe, L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin*, 21, 394-405.

[Bav95] Bavelas, J. (1995) Quantitative versus Qualitative? in *Social Approaches to Communication,* (W. Leeds-Hurwitz ed.), Guilford Press, London, UK.

[BCL+92] Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures, *Discourse Processes*, 15, 469-489

[BOO95] Bekker M, Olson J, and Olson M (1995), Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design, In: *Proc. of Symposium on Designing Interactive Systems (DIS 1995)*, ACM Press, New York.

[BD97] Bellotti V and Dourish P (1997) Rant and RAVE: experimental and experiential accounts of a media space. In: *Video Mediated Communication* (K. Finn, A. Sellen, and S. Wilbur eds.), Lawrence Erlbaum Associates, New Jersey.

[BBF+95] Benford S, Bowers J, Fahlen L, Greeenhalgh C, and Snowdon D (1995), User embodiment in collaborative virtual environments, In: *Proc. of the Conference on Human Factors in Computing Systems (CHI 1995)*, ACM Press, New York.

[BMB+09] Benko, H., Morris, M., Brush, A., and Wilson, A. (2009) Insights on Interactive Tabletops: A Survey of Researchers and Developers, Microsoft Research Technical Report, MSR-TR-2009-22, March 2009.

[BBB05] Berry L, Bartram L, Booth K (2005) Role-based control of shared application views, *Proc. UIST 2005*, Oct 23-26, 2005, Seattle, USA. ACM.

[BS97] Bertram, R. and Steinmetz, R., (1997) Scalability of audio quality for networked multimedia environments, *Proceedings of ICMCS'97*, Ottawa, Canada, pp. 294-301.

[BF91] Bier, E., and Freeman, S., (1991) MMM: A User Interface Architecture for Shared Editors on a Single Screen, *Proceedings of UIST '91*, Hilton Head, USA, November 11 – 13, 1991. ACM Press.

[BGM+07] Birnholtz, J., Grossman, T., Mak, C., and Balakrishnan, R. (2007) An exploratory study of input configuration and group process in a negotiation task using a large display, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA, April 28 - May 03, 2007). CHI '07. ACM, New York, NY

[BB03] Birnholtz J, Bietz M (2003) Data at work: supporting sharing in science and engineering, *Proc. GROUP 2003*, Sanibel Island, USA, Nov 9-12, 2003, ACM.

[Bly88] Bly, S., (1988) A Use of Drawing Surfaces in Different Collaborative Settings, *Proceedings of CSCW '88*, Portland, USA, 1988. ACM Press.

[BLM+06] Borwein J., Langstroth D., Macklem, M., Wilson, S., Jungic V., (2006) Coast-to-Coast Seminar and Remote Mathematical Collaboration, *Proc. HPCS 2006*, Saskatoon, Canada.

[BZO+08] Bos N, Zimmerman A, Olson J, Yew, J., Yerkle, J., Dahl, E., Conney, D., and Olson, G. (2007) From shared databases to communities of practice: A taxonomy of collaboratories, in , in *Scientific Collaboration on the Internet,* (G. Olson, A. Zimmerman, and N. Bos editors), MIT Press, Cambridge, Massachusetts.

[BZO+07] Bos N, Zimmerman A, Olson J, Yew, J., Yerkle, J., Dahl, E., Olson, G. (2007) From shared databases to communities of practice: A taxonomy of collaboratories. *J. of Comput.-Mediat. Commun.*, 12(2), 16. http://jcmc.indiana.edu/vol12/issue2/bos.html. Accessed June 1, 2008

[BSD00] Bouch, A., Sasse, M.A., and De Meer, H., (2000) Of packets and people: A user-centered approach to quality of service, *Proceedings of IWQoS 2000*, Pittsburgh, PA, June 5-8 2000.

[BS00] Bouch, A. and Sasse, M.A., (2000) The case for predictable media quality in networked multimedia applications, *Proceedings of ACM/SPIE Multimedia Computing and Networking (MMCN'00)*, 25-27 January 2000, San Jose, USA.

[BM85] Brinberg, D. and McGrath, J. (1985) *Validity and the Research Process,* SAGE Publication, Beverly Hills, USA.

[BM86] Buxton, B., and Myers. B., (1986) A study in two-handed input, *Proceedings of CHI '86*, Boston, USA, April 13 – 17, 1986. ACM Press.

[Bux92] Buxton, W. (1992) Telepresence: integrating shared task and person spaces. *Proceedings of Graphics Interface '92,* 123-129.

[BSS97] Buxton, B., Sellen, A., and Sheasby, M. (1997) Interfaces for Multiparty Videoconferences, in *Video Mediated Communication* (K. Finn, A. Sellen, and S. Wilbur editors), Lawrence Erlbaum and Associates, Mahwah, USA.

[Bux09] Buxton, B. (2009) Mediaspace – Meaningspace – Meetingspace, in *Media Space: 20+ Years of Mediated Life,* (S. Harrison, editor), Springer, 2009.

[Car08] Card, S. (2008) Information Visualization, *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications,* (A. Sears and J. Jacko, eds), Lawrence Erlbaum and Associates, New York.

[CMS99] Card, S., Mackinlay, J.D., and Shneiderman, B. (1999) Readings in information visualization: Using vision to think. Morgan Kauffman Publishers, Inc., San Francisco, CA

[Car03] Carlson J (2003) *iChat AV 2 for MAC OS X*, Peachpit Press, Berkley, CA, USA.

[Cer02] Cerrato L (2002) A coding scheme for feedback including gestures, Dept of Linguistics, Goteborg University, Sweden.
www.ling.gu.se/projekt/tal/doc/fb_with_gestures.pdf

[COP+00] Childers, L., T. Disz, R. Olson, M. E. Papka, R. Stevens, and T. Udeshi (2000) Access Grid: Immersive Group-to-Group Collaborative Visualization in *Proc. of Immersive Projection Technology Workshop*, (Ames, Iowa, USA, June 19-20, 2000).

[CCG+03] Chisan, J., Cockburn, J., Garner, R., Jazayeri, A., Kaminski, P., and Wesson, J. (2003) Video Bench, Computer Science 586a Final Report, Department of Computer Science, University of Victoria, www.ideanest.com/stuff/VideoBenchReport.pdf, April, 2003.

[CGR07] Church, R., Garber, P., and Rogalski, K. (2007) The role of gesture in memory and social communication, *Gesture,* 7 (2), 2007.

[Cla04] Clark, H. (1996), *Using Language,* Cambridge University Press, Cambridge, UK.

[CK04] Clark, H., & Krych, M. (2004) Speaking while monitoring addressees for understanding. *Journal of Memory & Language*, *50*(1), 62

[CZ09] Corrie, B., and Zimmerman, T. (2009) Build It: Will They Come? Media Spaces in the Support of Computational Science, in *Media Space: 20+ Years of Mediated Life* (ed. S. Harrison), Springer.

[CS07] Corrie B, Storey M. (2007) Towards understanding the importance of gesture in distributed scientific collaboration, *Int. J. of Knowl. Inf. Syst*., 13(2), Springer, London.

[CS05] Corrie, B. and Storey, M.A., (2005) Towards Understanding the Importance of Gesture in Distributed Collaborative Environments, *Workshop on Multimodal Interaction for the Visualization and Exploration of Scientific Data*, October 3, 2005, Trento, Italy.

[CZP+05] Corrie, B., Zimmerman, T., Patrick, A., El-Khatib, K., Singer, J., and Noel, S., (2005) Technology, Technology, Everywhere *Proceedings of the Workshop on Advanced Collaborative Environments,* WACE 2005, Sept. 8 - 9, 2005, Redmond, Washington.

[Cor03] Corrie, B., Co-located and Distributed Multi-User Interaction in a Digital Tabletop Environment, Technical Report DCS-291-IR, Department of Computer Science, University of Victoria, Victoria, Canada, 2003

[CFM+03] Corrie, B., Fung, J., Mueller, C., Smith, J., and Wong, J. (2003) VEGTables: Video Editing Gestures on Tabletops, Computer Science 544 Project Report, Department of Computer Science, University of British Columbia, December, 2003.

[CC05] Craft, B. and Cairns, P. (2005) Beyond Guidelines: What can we learn from the Information Seeking Mantra, *Proceedings of the Symposium on Information Visualization (INFOVIS 2005),* p. 110-118, IEEE Press.

[Cra09] Craven, G. (2009) *What's the Worst That Could Happen? A Rational Response to the Climate Change Debate,* Perigee Trade.

[Cre03] Creswell, J (2003), *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, Sage Publications, Thousand Oaks, CA.

[Cre94] Creswell, J (1994), *Research Design: Qualitative and Quantitative Approaches*, Sage Publications, Thousand Oaks, CA.

[CFH97] Cutler, L., Frolich, B., and Hanrahan, P., (1997) Two Handed Direct Manipulation on the Responsive Workbench, *Symposium of Interactive 3D Graphics*, Providence, USA, 1997, ACM Press.

[DL86] Daft, R. and Lengel, R. (1986) Organizational Information Requirements, Media Richness and Structural Design, *Management Science*, vol. 32, no. 5, pp. 554-571, 1986.

[DES+00] Damian, D., Eberlein, A., Shaw, M., and Gaines, B., Using Different Communication Media in Requirements Management, *IEEE Software*, 17(3), May 2000, pp. 28-36

[DFV08] Dennis, A., Fuller, R., and Valacich, S. (2008) Media, Tasks, and Communication Processes: A Theory of Media Synchronicity, *MIS Quarterly*, vol. 32, no. 3, pp. 575-560, Sept. 2008.

[Dev82] Devore, J. (1982) *Probability and Statistics for Engineering and the Sciences,* Brooks Cole, Monterey, California.

[DL01] Dietz, P., and Leigh, D., DiamondTouch: A multi-user Touch Technology, in *Proc. of UIST '01* (Orlando, USA, November 11 – 14, 2001) ACM Press.

[DB92] Dourish, P. and Bly, S. (1992) Portholes: supporting awareness in a distributed work group. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, United States, May 03 - 07, 1992). P. Bauersfeld, J. Bennett, and G. Lynch, Eds. CHI '92. ACM, New York, NY, 541-547.

[Egi88] Egido, C. (1988) Video conferencing as a technology to support group work: a review of its failures. In *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work* (Portland, Oregon, United States, September 26 - 28, 1988). CSCW '88. ACM, New York, NY

[EF94] Ellis, D and Fisher, B. (1994), *Small Group Decision Making: Communication and the Group Process*, McGraw Hill, New York

[EKL+03] Everitt, K. M., Klemmer, S. R., Lee, R., and Landay, J. A. (2003) Two worlds apart: bridging the gap between physical and virtual media for distributed design collaboration, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA, April 05 - 10, 2003). CHI '03. ACM, New York, NY, 553-560.

[Fin03] Finholt, T. (2003) Collaboratories as a New Form of Scientific Organization, *Journal of Econ. Innov. New Tech*. 12 (1), pp. 5 – 25, 2003, Routledge.

[FO97] Finholt, T., and Olson, G., (1997) From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychol. Sci.,* 8(1), Blackwell.

[FSW97] Finn, K., Sellen, A., and Wilbur, S. (1997) *Video Mediated Communication*, Lawrence Erlbaum Associates, Mahwah, NJ.

[FIB95] Fitzmaurice, G., Ishii, H., and Buxton, B., (1995) Bricks: Laying the Foundation for Graspable User Interfaces, *Proceedings of CHI '95*, Denver, USA, May 7 – 11, 1995. ACM Press.

[FBK+99] Fitzmaurice, G., Balakrishnan, R., Kurtenbach, G., and Buxton, B., (1999) An Exploration into Supporting Artwork Orientation in the User Interface, *Proceedings of CHI '99*, Pittsburgh, USA, May 15 – 20, 1999. ACM Press.

[FJH+00] Fox A, Johanson B, Hanrahan P, Winograd, T. (2000) Integrating information appliances into an interactive workspace, *IEEE CG&A,* 20(3), May/June 2000.

[Fur00] Furuyama, N. (2000) Gestural interaction between the instructor and the learner in origami instruction. *Language and Gesture,* (D. McNeill editor)), pp. 99-117. Cambridge : Cambridge University Press.

[FKS00] Fussell, S., Kraut, R., and Siegel, J., (2000) Coordination of communication: effects of shared visual context on collaborative work, in *Proc. of CSCW '00* (Philadelphia, Pennsylvania, USA, Dec. 2000) ACM Press.

[FSP03] Fussell, S., Setlock, L., and Parker, E. (2003) Where do helpers look?: gaze targets during collaborative physical tasks, *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, Ft. Lauderdale, Florida, USA, April 05 - 10, ACM Press.

[FSY+04] Fussell, S., Setlock, L., Yang, J., Ou, J., Mauer, E., and Kramer, A. (2004) Gestures over Video Stream to Support Remote Collaboration on Physical Tasks, *Human Computer Interaction,* 19 (3), pp. 273 – 309, Taylor and Francis.

[GMM+92] Gaver, W., Moran, T., MacLean, A., Lövstrand, L., Dourish, P., Carter, K., and Buxton, W. (1992) Realizing a video environment: EuroPARC's RAVE system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, United States, May 03 - 07, 1992). P. Bauersfeld, J. Bennett, and G. Lynch, Eds. CHI '92. ACM, New York, NY

[GSH+93] Gaver W, Sellen A, Heath C, and Luff P (1993) One is not enough: multiple views in a media space, in *Proc. of Conference on Human Factors in Computing Systems (CHI93)*, ACM Press, New York.

[GT02] Gherardi, S. and Turner, B. (2002) Real Men Don't Collect Soft Data, in *The Qualitative Researcher's Companion* (M. Huberman and M. Miles, editors.), Sage, Thousand Oaks, USA.

[GS67] Glaser, B. and Strauss, A. (1967) *The Discovery of Grounded Theory: strategies for qualitative research*, Aldine, Chicago.

[GDM09] Gorzynski, M., Derocher, M., and Mitchel, A. (2009) The Halo B2B Studio, in *Media Space: 20+ Years of Mediated Life* (S. Harrison editor), Springer.

[GG00] Gutwin, C., and Greenberg, S. The Mechanics of Collaboration: Developing Low Cost Usability Evaluation Methods for Shared Workspaces, in *Proc. of WETICE '00* (Gaithersburg, USA, March 14-16, 2000) IEEE Press.

[GG02] Gutwin, C., and Greenberg, S. (2002) A Descriptive Framework of Workspace Awareness for Real-Time Groupware, Computer Supported Collaborative Work, 11, 3-4, 411-446, Springer.

[GP02] Gutwin, C. and Penner, R. (2002) Improving interpretation of remote gestures with telepointer traces, in *Proc. of the 2002 ACM conference on Computer supported cooperative work (CSCW02)* (New Orleans, Louisiana, USA, November 16 - 20, 2002), ACM Press

[Ham62] Hamming R (1962) *Numerical Methods for Scientists and Engineers*, McGraw-Hill, New York.

[Har09] Harrison, S. (2009) *Media Space: 20+ Years of Mediated Life*, Springer, 2009.

[HBA+97] Harrison S, Bly S, Anderson S, Minneman S. (1997) The Media Space In Finn K, Sellen A, Wilbur S (eds) *Video Mediated Communication*, Lawrence Erlbaum Associates, Mahwah.

[HL91] Heath, C. and Luff, P. (1991) Disembodied conduct: communication through video in a multi-media office environment. In *Proceedings of the SIGCHI Conference on*

Human Factors in Computing Systems: Reaching Through Technology (New Orleans, Louisiana, United States, April 27 - May 02, 1991), ACM, New York, NY

[HL92] Heath, C. and Luff, P. (1992) Media Space and Communicative Asymmetries: Preliminary Observations of Video-Mediated Interaction, *Human-Computer Interaction*, 1992, Vol. 7, No. 3, Pages 315-346

[HLS95] Heath, C., Luff, P., and Sellen, A. (1995) Reconsidering the Virtual Workplace: Flexible Support for Collaborative Activity, *Proceedings of the 1995 Fourth European Conference on Computer Supported Cooperative Work (ECSCW '95)*, 11-15 September, 1995, Stockholm, Sweden

[HS92] Hollan, J. and Stornetta, S. (1992) Beyond being there, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, United States, May 03 - 07, 1992). ACM, New York, NY, 119-125

[Hue89] Huebcsh, J. (1989) *Communication 2000,* Butterworth, Durban, South Africa.

[HK03] Hunt, A. and Kingstone, A. (2003) Covert and overt voluntary attention: linked or independent? *Cognitive Brain Research,* 18, p. 102-105, Elsevier.

[IHH+10] Isenberg, H., Hinrichs, U., Hancock, H., and Carpendale, S. (2010) Digital Tables for Collaborative Information Eploration, *Tabletops – Horizontal Interactive Displays* (C. Muller-Tomfelde ed), Springer-Verlag, London.

[ITC08] Isenberg, P, Tang, A. and Carpendale, S. (2008) An Exploratory Study of Visual Information Analysis**,** *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pp. 1217-1226, ACM Press.

[Ish90] Ishii, H. (1990) TeamWorkStation: towards a seamless shared workspace. In *Proceedings of the 1990 ACM Conference on Computer-Supported Cooperative Work* (Los Angeles, California, United States, October 07 - 10, 1990), ACM, New York, NY, 13-26

[IK92] Ishii, H. and Kobayashi, M. (1992) ClearBoard: a seamless medium for shared drawing and conversation with eye contact, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, United States, May 03 - 07, 1992). ACM, New York, NY, 525-532

[IKG92] Ishii, H., Kobayashi, M., and Grudin, J. (1992) Integration of inter-personal space and shared workspace: ClearBoard design and experiments, *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work* (Toronto, Ontario, Canada, November 01 - 04, 1992), ACM, New York, NY, 33-42.

[iPhone] www.apple.com/iphone/

[IAC+07] Izadi, S., Agarawal, A., Criminisi, A., Winn, J., Blake, A., and Fitzgibbon, A. (2007) C-Slate: A Multi-Touch and Object Recognition System for Remote Collaboration using Horizontal Surfaces, IEEE Tabletop 07, Oct 10-12, IEEE Press.

[JMM+06] Johnson, C., Moorhead, R., Munzner, T., Pfister, H., Rheingans, P., and Yoo, T. (2006) *NIH/NSF Visualization Research Challenges Report*, IEEE Computer Society Press.

[Ken80] Kendon, A. (1980) Gesticulation and speech: Two aspects of the process of utterance, *The relation between verbal and nonverbal communication,* (M. Key editor), p. 207-227, Mouton, The Hague.

[Ken04] Kendon, A. (2004), *Gesture: Visible* Action *as Utterance*, Cambridge University Press, Cambridge, UK.

[KCR05] Kirk, D., Crabtree, A., and Rodden, T. (2005) Ways of the Hands, *Proceedings of ECSCW 2005*, pp. 3 – 10, Springer.

[KRF07] Kirk, D., Rodden, T., and Fraser, D. (2007) Turn it this way: grounding collaborative action with remote gestures, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, California, USA, April 28 - May 03, ACM Press.

[KDK99] Knoche, H., De Meer, H., and Kirsh, D., (1999) Utility curves: Mean Opinion Scores considered biased. *Proceedings of the 7th International Workshop on Quality of Service (IWQoS'99)*, London, England, June 1999.

[Knu07] Knudson, E. (2007) Fundamental Components of Attention, *Annual Review of Neuroscience,* 30, p. 57 – 78, Annual Reviews.

[KP90] Kraemer, K. and Pinsonneault, A. (1990) Technology and Groups: Assessment of the Empirical Research, in *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*, (J. Galegher, R. Kraut, and C. Egido eds), Lawrence Erlbaum and Associates, Hillsdale, New York.

[KGF02] Kraut, R., Gergle, D., and Fussell, S. (2002) The Use of Visual Information in Shared Virtual Spaces: Informing the Development of Virtual Co-Presence, in *Proc. of the 2002 ACM conference on Computer supported cooperative work (CSCW02) (*New Orleans, USA November 16-20, 2002) ACM Press.

[KMS96] Kraut, R., Miller, M., and Siegel, J., (1996) Collaboration in performance of physical tasks: effects on outcomes and communication, *Proc. of the 1996 ACM conference on Computer supported cooperative work (CSCW96),* Boston, USA, November 16-20, 1996) ACM Press

[KEG88] Kraut, R., Egido, C., and Galegher, J. (1988) Patterns of contact and communication in scientific research collaboration. In *Proceedings of the 1988 ACM*

Conference on Computer-Supported Cooperative Work (Portland, Oregon, United States, September 26 - 28, 1988). CSCW '88. ACM, New York, NY

[KW52] Kruskal, W. and Wallis, W. (1952) Use of ranks in one-criterion variance analysis, *Journal of the American Statistical Association* **47** (260): 583–621, December 1952.

[KFB+97] Kurtenbach, G., Fitzmaurice, G., Baudel, T., and Buxton, B., (1997) The Design of a GUI Paradigm based on Tablets, Two-hands, and Transparency, *Proceedings of CHI '97*, Atlanta, USA, March 22-27, 1997, ACM Press.

[KYY+99] Kuzuoka, H., Yamashita, J., Yamazaki, K., and Yamazaki, A. (1999) Agora: a remote collaboration system that enables mutual monitoring, *CHI '99 Extended Abstracts on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, May 15 - 20, 1999). CHI '99. ACM, New York, NY, 190-191.

[Las48] Laswell, H. (1948) The Structure and Function of Communication in Society, *The Communication of Ideas*, (L. Bryson editor), New York: Harper, 37–51.

[LRJ+06] Leigh, J., Renambot, L., Johnson, A., Jeong, B., Jagodic, R., Schwarz, N., Svistula, D., Singh, R., Aguilera, J., Wang, X., Vishwanath, V., Lopez, B., Sandin, D., Peterka, T., Girado, J., Kooima, R., Ge, J., Long, L., Verlo, A., DeFanti, T. A., Brown, M., Cox, D., Patterson, R., Dorn, P., Wefel, P., Levy, S., Talandis, J., Reitzer, J., Prudhomme, T., Coffin, T., Davis, B., Wielinga, P., Stolk, B., Koo, G. B., Kim, J., Han, S., Kim, J. W., Corrie, B., Zimmerman, T., Boulanger, P., and Garcia, M. (2006) The global lambda visualization facility: an international ultra-high-definition wide-area visualization collaboratory. *Future Gener. Comput. Syst.* 22, 8 (Oct. 2006)

[Lev60] Levene, H. (1960) Robust Tests for the Equality of Variance, *Contributions to Probability and Statistics*, (I. Olkin, eds.) Palo Alto, Calif.: Stanford University Press.

[LKH+09] Luff, P., Kuzuoka, H., Heath, C., Yamazaki, K., and Yamashita, J. (2009) Creating Assemblies in Media Space: Recent development in Enhancing Access to Workspaces, *Media Spaces: 20+ Years of Mediate Life,* (S. Harrison, editor), p. 27 – p.55, Springer.

[LHG92] Luff, P., Heath, C., and Greatbatch, D. (1992) Tasks-in-interaction: paper and screen based documentation in collaborative activity, *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work* (Toronto, Ontario, Canada, November 01 - 04, 1992). CSCW '92. ACM, New York, NY

[MW47] Mann, H., and Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other, *Annals of Mathematical Statistics, 18*, 50–60.

[MCK03] Mark, G., Carpenter, K., and Kobsa, A. (2003) A Model of Synchronous Collaborative Information Visualization, Proceedings of the International Conference on Information Visualization, July 16 – 18, London, England, IEEE Press.

[MAF95] Masoodian, M., Aooerkey, M., and Frederickson, L., (1995) Video support for shared work space interaction: an empirical study, *Interacting With Computers*, 7(3), 1995, pp. 237-253.

[MR97] Matsushita, N., and Rekimoto, J. (1997) HoloWal: Designing a Finger, Hand, Body and Object Sensitive Wall, *Proceedings of UIST '97*, Banff, Canada, 1997. ACM Press.

[MR06] Max, H. and Ray, T. (2006), *Skype: A Definitive Guide*, Que, Indianapolis, USA.

[Max02] Maxwell, J. (2002) Understanding and Validity in Qualitative Research, *The Qualitative Researcher's Companion* (M. Huberman and M. Miles editors), Sage Publications, Thousand Oaks, USA.

[Mil56] Miller, G.A. (1956) The Magical Number, Seven, Plus or Minus Two: Some limits on our capacity for processing information, *Psychological Review, 63* (1956). Pp. 81 – 97.

[MJ95] McCanne, S., and Jacobson, V. (1995) vic: A Flexible Framework Framework for Packet Video, *Proceedings of ACM Multimedia '95*, ACM Press.

[McG93] McGrath, J. (1993), A Typology of Tasks. *Groupware and Computer-Supported Cooperative Work* (R. Baecker, editor), Morgan Kaufmann: San Francisco, pp 165-168

[McG93b] McGrath, J. (1993), Methods for the Study of Groups, *Groupware and Computer-Supported Cooperative Work* (R. Baecker, editor), Morgan Kaufmann: San Francisco, pp 165-168

[McG84] McGrath, J. (1984) *Groups: Interaction and Performance*, Prentice Hall, 1984.

[McG91] McGrath, J. (1991) Time, Interaction, and Performance (TIP): A Theory of Groups, *Small Group Research,* 22 (2), May 1991, pp 147-174, Sage Publications.

[MPM+93] McKinlay, A., Procter, R,. Masting, O., Woodburn, R., and Arnott, J. (1993) A Study of Turn-Taking in a Computer-Supported Group Task, *Proc. of the Eighth Conference of the British Computer Society Human Computer Interaction Specialist Group* (Loughborough University, UK August 7-10, 1993)

[McN85] McNeill, D. (1985) So you think gestures are nonverbal, *Psychological Review,* Volume 92, p. 350-371.

[McN92] McNeill, D. (1992) *Hand and Mind: What Gestures Reveal about Thought*, University of Chicago Press, Chicago, USA.

[Merriam] Merriam-Webster Online Dictionary (2010) Merriam-Webster Online. 24 January 2010, www.merriam-webster.com/dictionary

[MI04] Miwa, Y. and Ishibiki, C. (2004) Shadow communication: system for embodied interaction with remote partners. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (Chicago, Illinois, USA, November 06 - 10, 2004). CSCW '04. ACM, New York, NY

[MKM+09] Mol, L., Krahmer, E., Maes, A., and Swerts, M. (2009) The communicative import of gestures: Evidence from a comparative analysis of human-human and human-machine interactions, *Gesture,* 9 (1), John Benjamins.

[MJA+02] Mullin J, Jackson M, Anderson A, Smallwood L, Sasse A, Watson A, and Wilson G (2002) The ETNA Taxonomy, www-mice.cs.ucl.ac.uk/multimedia/projects/etna/taxonomy.pdf

[NSK+93] Nardi B, Schwarz H, Kuchinsky A, Leichner R, Whittaker S, and Sclabassi R (1993) Turning away from talking heads: the use of video-as-data in neurosurgery, *Proc. of the Conference on Human Factors in Computing Systems (CHI 1993)*, ACM Press, New York.

[NRC93] National Research Council (U.S.) (1993) National Collaboratories: Applying Information Technology for Scientific Research. Washington, D.C: National Academy Press.

[NTC07] Neumann, P., Tang, A. and Carpendale, S. (2007) A Framework for Visual Information Analysis. Research report 2007-87123, University of Calgary, Calgary, AB, Canada, July, 2007.

[Nou04] Norusis, M. (2004) *SPSS 13.0 Guide to Data Analysis*. Upper Saddle-River, N.J.: Prentice Hall, Inc.

[OZB08] Olson, G., Zimmerman, A., and Bos, N. (2008) *Scientific Collaboration on the Internet*, MIT Press, Cambridge, Massachusetts.

[OHB+08] Olson, J., Hofer, E., Bos, N., Zimmerman, A., Olson, G., Cooney, D., and Faniel, I. (2008) A Theory of Remote Scientific Collaboration, in *Scientific Collaboration on the Internet,* (G. Olson, A. Zimmerman, and N. Bos editors), MIT Press, Cambridge, Massachusetts.

[OEJ+08] Olson, J., Ellisman, M., James, M., Grethe, J., and Puetz, M. (2008) The Biomedical Informatix Research Network, in *Scientific Collaboration on the Internet,* (G. Olson, A. Zimmerman, and N. Bos editors), MIT Press, Cambridge, Massachusetts.

[OTC+02] Olson, J., Teasley, S., Covi, L., and Olson, G. (2002) The (Currently) Unique Advantages of Collocated Work, in *Distributed Work* (P. Hinds and S. Keisler, editors), MIT Press, Cambridge Massachusetts

[OO01] Olson, G. and Olson, J. (2001), Technology Support for Collaborative Work Groups, in *Coordination Theory and Collaboration Technology,* (G. Olson ed), Lawrence Erlbaum Associates, Mahwah, NJ, USA, pp. 559 – 584.

[OO00] Olson G, Olson J (2000) Distance Matters, *J. Hum.-Comp. Interact*., 15(2/3), Lawrence Erlbaum Associates, Mahwah.

[OO97] Olson, G. and Olson, J. (1997) Making Sense of the Findings: Common Vocabulary Leads to Synthesis Necessary for Theory Building, in *Video Mediated Communication*, K. Finn, A. Sellen, and S. Wilbur eds., Lawrence Erlbaum Associates, Mahwah, New Jersey.

[OOM95] Olson J, Olson G, and Meader D (1995), What mix of video and audio is useful for small groups doing remote real-time design work? *Proc. of the Conference on Human Factors in Computing Systems (CHI 1995)*, ACM Press, New York.

[OFC+03] Ou, J., Fussell, S., Chen, X., Setlock, L., and Yang, J., (2003) Gestural Communication over Video Stream: Supporting Multimodal Interaction for Remote Collaborative Physical Tasks, *Proc. of the 5th international conference on Multimodal interfaces (ICMI 03)* (Vancouver, Canada, November 5-7, 2003) ACM Press.

[OSW+06] Ou J, Shi Y, Wong J, Fussel S, and Yang J (2006), Combining audio and video to predict helpers' focus of attention in multiparty remote collaboration on physical tasks, *Proc. of the 8th International Conference on Multimodal Interfaces (ICMI 2006),* ACM Press, New York.

[Ozy02] Ozyurek, A. (2002) Do Speakers Design their Cospeech Gestures for their Addressees? The Effects of Addressee Location on Representational Gestures, *Journal of Memory and Langauge,* 46, p. 688-704, Academic Press.

[PS99] Parks, C. and Sanna, L. (1999) *Group Performance and Interaction,* WestView Press, Oxford, UK.

[PKL00] Park, K., Kapoor, A., and Leigh, J. (2000) Lessons Learned from Employing Multiple Perspectives In a Collaborative Virtual Environment for Visualizing Scientific Data. *Proc. of the Conference on Collaborative Virtual Environments (CVE),* pp. 73–82, New York, USA, ACM Press.

[PSC+04] Patrick, A.S., Singer, J., Corrie, B., Noël, S., El Khatib, K., Emond, B., Zimmerman, T., & Marsh, S. (2004). A QoE Sensitive Architecture for Advanced Collaborative Environments. *First International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QSHINE 2004),* Oct. 18-20, Dallas, TX.

[PIH01] Patten, J. Ishii, H., Hines, J., and Pangaro, G., (2001) Sensetable: A Wireless Object Tracking Platform for Tangible User Interfaces, *Proceedings of CHI '01*, Seattle, USA, 2001. ACM Press.

[PMM+93] Pederson, E., McCall, K., Moran, T., and Halasz, F. (1993) Tivoli: An Electronic Whiteboard for Informal Workgroup Meetings, *Proceedings of CHI '93*, Amsterdam, The Netherlands, April 24-29, 1993. ACM Press.

[PBN09] Pinelle, D., Barjawi, M., Nacenta, M., and Mandryk, R. (2009) An evaluation of coordination techniques for protecting objects and territories in tabletop groupware. *Proceedings of the Conference on Human Factors in Computing Systems*, Boston, MA, USA, April 04 - 09, 2009, ACM, New York.

[Quintilian] *Institutes of Oratory*. Ed. Lee Honeycutt. Trans. John Selby Watson. 2006. Iowa State University, Downloaded Dec. 2, 2009. <http://honeyl.public.iastate.edu/quintilian/>

[Rek98] Rekimoto, J., (1998) A Multiple Device Approach for Supporting Whiteboard-based Interactions, *Proceedings of CHI '98*, Los Angeles, USA, April 18-23, 1998. ACM Press.

[RSW+98] Richardson T, Stafford-Fraser Q, Wood K, and Hopper A (1998), Virtual Network Computing, IEEE Internet Computing, Vol. 2, No. 1, 33-38.

[RHV98] Rimell, A., Hollier, M., and Voelcker, R., The influence of cross-modal interaction on audio-visual speech quality perception, Presented at the 105th Convention AES. September 26-29, San Francisco. *Audio Engineering Society Preprint 4791*.

[RB87] Ruesch, J. Bateson, G. (1987) *Communication: The Social Matrix of Psychiatry*, Norton, W. W. & Company, 1987.

[Sch54] Schramm, W (1954) How communication works, *The Process and Effects of Mass Communication,* Schramm W (ed.), Urbana: University of Illinois Press

[SC10] Scott, S. and Carpendale, S. (2010) Theory of Tabletop Territoriality, *Tabletops – Horizontal Interactive Displays* (C. Muller-Tomfelde editor), p. 357, Springer-Verlag, London.

[SGM03] Scott, S., Grant, K., and Mandryk, R. (2003). System Guidelines for Co-located, Collaborative Work on a Tabletop Display, *Proc. of European Conference on Computer Supported Cooperative Work (ECSW 2003)*, pg. 159-178.

[Sel92] Sellen, A. J. (1992) Speech patterns in video-mediated conversations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, United States, May 03 - 07, 1992), ACM, New York, NY, 49-59

[SW49] Shannon, C. and Weaver, W. (1949) *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Illinois.

[SLV+02] Shen, C., Lesh, N., Vernier, F., Forlines, C., and Frost, J., (2002) Sharing and Building Digital Group Histories, *Proceedings of CSCW '02,* New Orleans, USA, November 16 – 20, 2002. ACM Press.

[Shn96] Shneiderman, B. (1996) The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization, *Proceedings of the Symposium on Visual Languages*, p. 336-343, Los Alamitos, USA, IEEE Press.

[Smart] www.smarttech.com

[SWM08] Sonnenwald, D.H., Whitton, M.C. & Magluaghlin, K. (2008). Evaluation of a scientific collaboratory system: Investigating a collaboratory's potential before deployment, *Scientific Collaboration on the Internet*, (G. Olson, A. Zimmerman & N. Bos Eds.), pp.171-194, MIT Press, Boston.

[SD06] Spivey, M. and Dale, R. (2006) Continuous Dynamics in Real-Time Cognition, *Current Directions in Psychological Science*, Vol. 15 Issue 5, pp. 207-211.

[SWM03] Sonnenwald D, Whitton M, Maglaughlin K. (2003) Evaluating a scientific collaboratory: Results of a controlled experiment. *ACM Trans. Comput.-Hum. Interact.* 10(2) 2003.

[SWS+02] Stahl, O., Wallberg, A., Soderberg, J., Humble, J., Fahlen, L., Bullock, A., and Lundberg, J. (2002) Information Exploration Using The Pond, *Proceedings of CVE 02*, Bonn, Germany, September 30 – October 2, 2002. ACM. Press

[SBD99] Stewart, J., Bederson, B., and Druin, A., (1999) Single Display Groupware: A Model for Co-present Collaboration, *Proceedings of CHI '99*, Pittsburgh, USA, May 15-20, 1999. ACM Press.

[SGH+99] Streitz, N., Geibler, J., Holmer, T., Konomi, S., Muller-Tomfelde, C., Reischl, W., Rexforth, P., Seitz, P., and Steinmetz, R., (1999) I-LAND: An interactive Landscape for Creativity and Innovation, *Proceedings of CHI '99*, Pittsburgh, USA, May 15-20, 1999. ACM Press.

[SPM+02] Streitz, N., Prante, T., Muller-Tomfelde, C., Tandler, P., and Magerkurth, C. (2002) Roomware – The Second Generation, *CHI '02 Extended Abstracts*, Minneapolis, USA, April 20-25, 2002. ACM Press.

[SP98] Strickon, J., and Paradiso, J., (1998) Tracking Hands above Large Interactive Surfaces with a Low-Cost Scanning Laser Rangefinder, *Conference Summary of CHI '98*, Los Angeles, USA, April 18-23. ACM Press.

[Stu09] Stults, R. (2009) Media Space After 20 Years, *Media Space: 20+ Years of Mediated Life,* (S. Harrison, editor), Springer, 2009.

[Swe88] Sweller, J., Cognitive Load During Problem Solving: Effects on Learning, *Cognitive Science*, 12 (2), April-June 1988, pp. 257-285.

[HT87] Hochberg, Y., and Tamhane, A. (1987) *Multiple Comparison Procedures*, Wiley series in Probability and Mathematical Statistics, pp 193-194.

[Tandberg09] Tandberg Question and Answers: Video Communications Industry Backgrounder, Andrew Davis, Wainhouse Research, Downloaded Nov 23, 2009. http://www.tandberg.com/collateral/Video_Communications_Information_Backgrounder.pdf

[TSE+95] Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., and Sedivy, J. (1995) Integration of Visual and Linguistic Information in Spoken Language Comprehension, *Science,* 268 (5217), pp. 1632-1634.

[TNG04] Tang, A., Neustaedter, C. and Greenberg, S. (2004) Embodiments and VideoArms in Mixed Presence Groupware**,** Report 2004-741-06, Department of Computer Science, University of Calgary, Calgary, Alberta, Canada March 2004.

[TNG06] Tang, A**.**, Neustaedter, C., and Greenberg, S. (2006). VideoArms: Embodiments for Mixed Presence Groupware. In Proceedings of the 20th British HCI Group Annual Conference (HCI 2006). (September 11-15, Queen Mary, University of London). pp: 85-102.

[TPI+10] Tang, A., Pahud, M., Inkpen, K., Benko, H., Tang, J., and Buxton, B. (2010) Three's Company: Understanding Collaboration Channels in Three-Way Distribute Collaboration, In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Savannah, GA, USA, February 6 - 10, 2010), ACM, New York, NY

[TM91] Tang, J. C. and Minneman, S. (1991) VideoWhiteboard: video shadows to support remote collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching Through Technology* (New Orleans, Louisiana, United States, April 27 - May 02, 1991), ACM, New York, NY, 315-322

[Tan89] Tang, J. (1989) Listing, Drawing, and Geturing in Design: A study of the Use of Shared Workspace by Design Teams, Xerox PARC Technical Report, SSL-89-3.

[TM90] Tang, J. and Minneman, S., (1990) VideoDraw: a video interface for collaborative drawing, *Proc. of the SIGCHI conference on Human factors in computing systems (CHI90)* (Seattle, USA, April 01 - 05, 1990), ACM Press.

[TL88] Tang J and Leifer L (1988), A Framework for Understanding the Workspace Activity of Design Teams, *Proc. of the Conference on Computer Supported Cooperative Work (CSCW 1988)*, ACM Press, New York.

[TCS97] Tindall-Ford, S., Chandler, P., and Sweller, J. (1997) When Two Sensory Modes Are Better Than One, *Journal of Experimental Psychology,* 3 (4), p. 257 – 287.

[TM04] Tory, M. and Moller, T. (2004) Rethinking Visualization: A High-Level Taxonomy, *Proceedings of the Symposium on Information Visualization (INFOVIS 2004),* p. 151 – 158, Austin, Texas, IEEE Press.

[TSG+06] Tse, E., Shen, C., Greenberg, S., and Forlines, C. (2006) Enabling interaction with single user applications through speech and gestures on a multi-user tabletop, *Proceedings of the Working Conference on Advanced Visual interfaces* (Venezia, Italy, May 23 - 26, 2006). AVI '06. ACM, New York, NY, 336-343.

[TR09] Tuddenham, P. and Robinson, P. (2009) Territorial coordination and workspace awareness in remote tabletop collaboration, *Proceedings of the 27th international*

Conference on Human Factors in Computing Systems, Boston, MA, USA, April 04 - 09, 2009), ACM, New York.

[TR10] Tuddenham, P. and Robinson, P. (2010) Coordination and Awareness in Remote Tabletop Collaboration, *Tabletops – Horizontal Interactive Displays* (C. Muller-Tomfelde editor), p. 407, Springer-Verlag, London.

[UCL] UCL Network and Multimedia Software. http://wwwmice.cs.ucl.ac.uk/multimedia

[UI97] Ullmer, B., and Ishii, H., (1997) MetaDESK: Models and Prototypes for Tangible User Interfaces, *Proceedings of UIST '97*, Banff, Canada, October 14-17, 1997. ACM Press

[UFK89] Upson, C., Faulhaber, T., Kamins, D., Laidlaw, D., Schlegel, D., Vroom, J., Gurwitz, R., and van Dam, A. (1989) "The Application Visualization System: A Computational Environment for Scientific Visualization," IEEE Computer Graphics and Applications, 9 (4), pp. 30-42.

[VOO+99] Veinott, E., Olson, J., Olson, G., and Fu, X., (1999) Video helps remote work: speakers who need to negotiate common ground benefit from seeing each other, *Proceedings of CHI '99*, May 15-20, 1999, Pittsburgh, Pennsylvania. ACM Press.

[VTC+10] Venolia, G., Tang, J., Cervantes, R., Bly, S., Robertson, G., Lee, B., and Inkpen, K. (2010) Embodied Social Proxy: Mediating Interpersonal Connection in Hub-and-Satellite Teams, *Proceedings of CHI 2010*, ACM Press.

[VLS02] Vernier, F., Lesh, N., and Shen, C., (2002) Visualization techniques for circular tabletop interfaces, *ACM Conference on Advanced Visual Interfaces,* Also available as Mitsubishi Engineering Research Lab Technical Report MERL TR 2002-01.

[Wae89] Waern, Y. (1989) *Cognitive Aspects of Computer Supported Tasks*, John Wiley and Sons, Toronto.

[War04] Ware, C. (2004) *Information Visualization: Perception for Design*, Morgan Kaufman, San Francisco.

[WS97] Watson, A. and Sasse, M., (1997) Multimedia conferencing via multicast: Determining the quality of service required by the end user, *Proceedings of AVSPN'97 - International Workshop on Audio-Visual Services over Packet Networks*, Aberdeen, Scotland, 15-16 September 1997.

[Wat01] Watson, A. (2001) Assessing the Quality of Audio and Video Components in Desktop Multimedia Conferencing, PhD Dissertation, Department of Computer Science, University College, London, March 2001.

[Wel93] Wellner, P., (1993) Interacting with Paper on the Digital Desk, *Communications of the ACM*, 36 (7), July 1993, ACM Press.

[WC97] Whittaker, S. and O'Conaill, B. (1997) The Role of Vision in Face-to-Face and Mediated Communication, *Video Mediated Communication* (K. Finn, A. Sellen, and S. Wilbur editors), Lawrence Erlbaum and Associates, Mahwah, USA.

[WJF+09] Wigdor D, Jiang H, Forlines C, Borkin M, Shen C (2009) WeSpace: The design development and deployment of a walk-up and share multi-surface visual collaboration system, *Proceedings of human factors in computing systems (CHI '09),* pp. 1237–1246, ACM Press, New York.

[WSF+07] Wigdor, D., Shen, C., Forlines, C., and Balakrishnan, R. (2007) Perception of elementary graphical elements in tabletop and multi-surface environments, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, California, April 28 - May 03, 2007, ACM, New York.

[Wil77] Williams, E. (1977) Experimental comparisons of face-to-face and mediated communication: a review, *Psychological Bulletin*, 84, 963-976.

[Wul89] Wulf, W. (1989), The National Collaboratory—A White Paper, Appendix A in *Towards a National Collaboratory*, the unpublished report of an invitational workshop held at the Rockefeller University, March 17-18, 1989 (Joshua Lederberg and Keith Uncapher, co-chairs).

[ZB98] Zigurs, I. and Buckland, B. (1998) A Theory of Task/Technology Fit and Group Support Systems Effectiveness, *MIS Quarterly*, vol. 22, no. 3, pp. 313-334, Sep 1998.

15 Appendices

15.1 Gesture Study: Limitations

15.1.1 Limitations of studying one-way communication

In this appendix we consider the limitations of studying one-way communication as it is used in out gesture study (Chapter 8 through 11). Two-way interaction between subject and presenter is removed from the study as part of the design, as it allows the study to focus on the question under consideration, that of how do subjects attend to gesture, and does that gesture help understanding? We consider this restriction on several dimensions.

First, our CoGScience task analysis (Section 8.1.1 and Appendix 15.6) and our observations of research presentations during research meetings and research seminars in Chapter 6 and Chapter 7 demonstrates that remote presentations are often consist of one way communication. For example, in Section 7.3.1 we see that many research meetings have long periods of one-way communication when a presenter is talking about a specific topic. Although limiting subjects so that they cannot interact with the presenter limits the generalizability of the results to more interactive collaboration tasks, it is a reasonable restriction to make for a presentation task. This study simulates such interactions quite effectively.

Second, the recorded part of the presentation does have interaction between the presenter and the Devil's Advocate who questions the presenter. This simulates a distributed interaction between two users, to which the subject is a passive participant. The video typically shows the presenter, but when the Devil's Advocate asks questions the video switches to video of the Devil's Advocate. This simulates "Voice Activated Switching" (VAS), a technique that is common among many H323 video conferencing systems. VAS switching means that the video of the individual who has the highest audio levels is displayed (the remote participants see whoever is talking).

Third, a study format in which study participants are non-interactive observers of a face-to-face dialogue is widely used by social psychologists to study face-to-face interaction, and in particular interaction that involves gesture [BC06]. Utilizing a one-way communication to study the decoding process of gestural interaction helps manage the complexities of the face-to-face communication process [BC00][BC06]. In particular, Bavelas and Chovil point out that although this design has its limitations, it is almost

exclusively used for studying the decoding process in gestural communication [BC06]. They also point out the two main limitations to this type of study. First, in many such studies the entire dialogue between the participants is not presented to the subjects. Since we include the entire presentation, simulating a VAS environment to provide video of both the presentation and the questions, the entire dialogue is provided to each subject. The second limitation is that, although subjects observe the entire communication, they are not active participants in the conversation. It has been shown that subjects who do not actively participate in dialogue have a poorer understanding of the dialogue, and Bavelas and Chovil suggest this might be true of gestural communication as well [BC06]. Although we are unaware of any studies that show this is true, it is important that we recognize this as an issue that can impact the generalizability of this study. As with any study design, there are limitations to the method used here. At the same time, the benefits of controlling the large number of variables with the passive observer approach makes this class of study extremely useful for the concise study of the decoding process in scientific communication.

Lastly, the experimental apparatus we use to determine what subjects are attending to (an eye-tracker that gathers highly accurate gaze data) is limited in the amount of participant movement that it allows. Thus studying a highly interactive communication task, where the subjects move around a lot, would be impossible using this technology. The ability to utilize eye-tracked gaze data to study gestural interaction in this manner is, to our knowledge, unique to this study.

15.1.2 Limitations of the Tobii tracking system

There are several limitations of the Tobii system that caused problems during this study. Firstly, although the eye-tracking system can be calibrated successfully for most subjects (including subjects with glasses), it is not possible to calibrate for some subjects. In particular, the calibration process fails for subjects who have glasses with a very strong prescription, subjects with stigmatisms, or subjects with a “lazy” eye (eyes that do not focus on the same location at the same time). In these cases, subjects were rejected before the study began for that subject and the next subject is assigned to that condition. High quality calibrations were achieved for all subjects used in this study.

When the Tobii system is not able to track an individual (when they blink, look off screen, etc.) the tracker reports that it is not tracking. For example, the box with the two circles in Figure 44 (p. 209) would show zero or one circle and the green box in the figure would be red or orange. When tacking is lost it is also indicated in the analysis data that the system exports. This is not an issue, as our analysis takes this occurrence into consideration.

More problematic are tracking inaccuracies and inconsistencies that are experienced in the midst of the study. The problem arises when the calibration works, the system reports that it is tracking normally, but there are errors in the tracking data provided from the system. Although the Tobii system specification says that it can track effectively when subjects move around within an acceptable region (a 26 x 20 x 32 cm area at distance of 73 cm), we experienced a problem in some cases. During pre-trial testing, it was found that certain types of movements introduced an offset in the gaze measurements that the system recorded. This was extremely problematic in that the system was reporting that it was tracking fine but it was clearly not tracking what was being looked at accurately.

This was clearly demonstrated in pre-trial testing. In some cases, test subjects were asked to look at specific parts of the screen during the video (e.g. the presenter's nose). When this data was analyzed the tracking system was reporting that the subjects were looking above or below the target. What makes this problematic is that the error offset does not exist when a subject starts watching the video, but part way through the video a very clear offset becomes present and is often large enough and consistent enough to be noticeable by the experimenter using the naked eye. We were unable to determine a specific type of movement that caused this error. In addition, it was not possible to reproduce the error even when testing across the same pre-trial subjects. This issue was raised with the manufacturer and no resolution could be found to the problem.

In order to detect this situation, a calibration section (similar to the Tobii calibration) was added to the end of the video (Act 8). Subjects were asked to look at a circle on the screen for two seconds. This was repeated five times, with the circle appearing at a different location on the screen each time. If significant errors were detected in this Act, we know that this subject's data is suspect. Fortunately, the post calibration act was very effective at identifying subjects where this problem occurred. Subject eye fixations that

occurred within the AOI regions that contained the circle were analyzed. Subjects that did not have eye fixation in more than two of the five AOI regions were rejected. Borderline subjects that had fixations in three of the five AOI regions were rejected if the total fixation time in the AOI regions was less than 20% of the total calibration scene duration time. Note that in subjects that were not rejected from the study the average percentage of fixation time in the calibration AOIs is 46%. There were twelve subjects that were rejected because they did not have eye fixations in more than two calibration AOIs. One subject was rejected on our less stringent criteria (fixations in three calibration AOIs and an AOI fixation percentage of 18%). One subject was accepted on this criterion (fixations in three calibration AOIs and an AOI fixation percentage of 44%). All other subjects were accepted with fixations in either four or five of the calibration AOIs. This left thirty-seven subjects in the gaze analysis portion of the study. Note that the questionnaire data was used for all fifty subjects, and only the gaze related data was discarded.

15.2 VideoBench: The Video Bench Application

As described in Section 1.1.1, the focus of this research changed from artifact-centric collaboration in general to the specific domain of artifact-centric, scientific collaboration *after* our initial experiments with the CoTable system. Our experiences with CoTable did not revolve around scientific collaboration, but instead considered another common artifact-centric task – that of video editing. Video editing was chosen as an artifact-centric task on CoTable for two primary reasons. First, it is a visually complex artifact-centric task (manipulating and editing video segments) and requires high fidelity control (control is required at frame level for editing videos). Second, video editing, like photo-editing [SLV+02], is a social activity that is conducive to tabletop interaction. As digital media (photos, music, and videos) becomes more prominent, many researchers predict that advanced technologies and applications for interacting with such media in a social context will become widespread [BMB+09]. Thus, video editing seemed like an excellent artifact-centric collaboration task to consider in this context.

Our implementation of a tabletop video editing application is called VideoBench, an interactive application for the manipulation and editing of video clips. The initial collocated application was developed by a group of students at the University of Victoria to explore gestural interaction on a tabletop device [CCG+03]. The application supported

collocated collaboration and used gesture as an interaction mechanism. It allows the user to play (play, pause, rewind, fast forward, stop) and edit (cut, splice, resize) a set of video clips using gestures on DT table. VideoBench is a relatively complex Java program, utilizing sophisticated imaging, media, and advanced interaction techniques (including gesture recognition on the DT table). The original application consisted of over 60 classes and made use of two complex Java packages, the piccolo imaging package and the Java Media Framework (JMF). The user interacts with the application through either mouse interaction or touch interaction on the DT. In both cases, user input (mouse clicks, gestures) is translated into actions on the media objects in the system (play, stop, splice, etc.). An image of the VideoBench application being used can be seen in Figure 9.

The original VideoBench application was extended to create an abstraction between input event and actions on the media objects. Mouse and gesture event handling no longer manipulate media objects directly, but instead are translated into abstract actions (CommandPlay for example). The CommandPlay abstraction knows how to make a media object play, and is independent of how the user indicates that the media object should start playing. Thus the play command can originate from a mouse click, a gesture, or even a voice command (not supported but within the capabilities of the abstraction).

15.2.1 Gestures in VideoBench

There are a number of gestures that are implemented as actions in the VideoBench application. Note that these are HCI gestures (communicating to the computer) rather than HHI gestures (communicating to a person). Users can “cut” a video clip into two smaller clips by making a slicing gesture across the video clip. Users can advance or rewind the video clip by making a circular motion on the video clip (like the “jog” dial on a DVD player). Video clips are moved by touching and dragging them. The user can move two video clips at the same time. Dragging two video clips so that their edges touch joins the clips together. Fast-forward and rewind commands are issued by touching a video clip and rotating their fingers clockwise or counter clockwise much like a jog-dial on professional video editing hardware.

It is worth noting that our primary interest in this research is in human-to-human interaction gestures (HHI). That is, we are exploring how people communicate with each other at a distance using gestures, not the HCI aspect of how they interact with the

VideoBench software using the currently implemented gesture set. The gesture set that is supported is important, because it enables natural, gesture based interaction with digital artifacts, but it is not the focus of the research presented here. Although we have carried out a study of how gestures were used in the collocated version of VideoBench [CFM+03] that research is not discussed in this dissertation.

15.2.2 Distributed VideoBench

We extended the original face-to-face VideoBench application to create a new distributed VideoBench application. The distributed application functions like the collocated version except that there are now applications running at two or more physically disjoint locations. The applications exchange state such that when a user at one site performs an operation the user at the remote site sees the operation occur as if it was performed locally.

We utilize the command abstraction described in Section 5.1.3 to "send" interaction events within the application, from user interaction to the media-processing engine. The main reason for using this design is the ability to incorporate sending application events such as CommandPlay to applications that are distributed over the network. Thus the application not only sends interaction events to the application itself, but it can also send events to distant applications. Thus when a command is issued on one device (either mouse or DT controlled), the command is sent over the network to other connected devices. The remote devices receive the commands and issue them to the local application as if they are generated locally. Thus in some ways it appears that the remote users are interacting with the local application.

The user's gestures are communicated to the user at the remote site through the drawing of an icon (in this case, a circle of roughly fingertip size) that represents where the remote user is touching the table. Up to two contact points can be communicated per user. In order to communicate user actions, each touch point leaves behind a trace as the user interacts with the table. This gestural trace gradually fades out over a short period (approximately 1 second). Such gestural traces have been shown to be important in helping users communicate effectively when network performance (in particular, variable latency in the network) is an issue [GP02]. Although our system does not suffer from variable latency problems (see below for details), such gesture traces also appear to

provide a more coherent reference space for collaborators. As Buxton suggests (without quantitative support, mind you), a simple mouse pointer as a mechanism for distributed pointing provides the *"… gestural vocabulary of a fruit fly"* [Bux09]. With gestural traces, we hope to be able to communicate the gestural vocabulary of at least a house fly.

All components in the CoTable and VideoBench systems are connected using standard IP networking over a 100 Mbps Ethernet network. All video, audio, gesture, and application state that is communicated between the remote collaborators is communicated over this network. The aggregate bandwidth required for this collaborative application is less than 10 Mbps. More details of the implementation of the distributed application can be found in [Cor03].

15.2.3 Technical issues with VideoBench

We encountered a number of technical issues in the development of the VideoBench system. The original authors of the VideoBench software reported issues in the accuracy and reliability of the DT touch detection [CCG+03]. We experienced similar issues. Accuracy is not as high as desired, with a level of variability in the reported position of a touch even when the touch is not moving. In addition, the DT system reports sporadic "no-touch" events when a user is touching the table. This makes the recognition of gestures very difficult and requires significant processing of the input to give smooth gesture recognition. Unfortunately it was necessary to change the way gesture processing was carried out from the original VideoBench system. This is primarily attributed to differences in the tables used (different versions of the DT table) and processor speed of the computer used to control the DT table. To make the application work on a wide variety of devices, it was decided that the application should err on the side of smooth functionality rather than give erroneous no-touch events. This unfortunately increases the latency in recognizing gestures but provides the benefit of more robust gestures.

Chisan *et al.* [CCG+03] also reported that multi-touch (multiple fingers) processing is not very robust on the DiamondTouch, requiring heuristics to correctly disambiguate multi-touch interaction correctly. The implemented heuristics in the VideoBench application often do not perform adequately, sometimes reversing the contact points on the bounding box of the contact area. In addition, the processing of the second touch suffers from the same "no-touch" problem as a single touch. The processing required to

make both of these functionalities more robust is possible, but was considered outside the scope of the work we were carrying out. Improvements for a single touch were made but multi-touch was not. Thus multi-touch gestures did not perform well and often result in errors.

There were also some state consistency problems in the distributed VideoBench application. In particular, operations check to see if a video clip is "busy" performing another operation before executing the operation, and if so, a message is displayed and the operation is not performed. Since a distributed locking mechanism was not used to synchronize "busy" clips, it is possible for operations to erroneously succeed on a video on a local machine. Consider the following scenario. Machine A executes a play operation, and sends that information to machine B. At the same time, machine B executes a split operation on the same video. If machine B executes the split operation before the play command arrives, an inconsistent state results (since the video as a single entity no longer exists on machine B). This problem could be solved using distributed locking algorithms, but for the purposes of the experiments carried out here, this was not done. Although when this occurred it was confusing to the users, it did not occur often.

15.3 Ethnography: Focus Group Script

Welcome

Hello everyone, and thanks for participating in the focus group today. The purpose of this meeting is to explore, in some depth, how you use collaboration in your research. The goal is to do this through discussion amongst yourself based around your experiences. My role here is primarily as the facilitator, and I will not take active part in the discussion other than to keep it focussed and moving along. To this end, I have prepared a set of questions that will hopefully spark discussion and at the same time help keep the discussion moving along.

A brief reminder about the context for this focus group… The focus group is part of the study that I have been carrying out with your group over the past five months. Recall that what is said in this meeting is confidential and that none of you will be identified as participating in this study. Also, recall that if you want to withdraw from participation at any time you should feel free to do so. I am recording the meeting so I can explore your responses in more detail later but this video will only be used for analysis purposes.

Does anyone have any questions at this time???

Review of agenda

0:00 – 0:05 Welcome
0:05 – 0:10 Review of Agenda
0:10 – 0:40 Questions and discussion
0:40 – 0:50 Break
0:50 – 1:20 Questions and discussion
1:20 – 1:30 Wrap up

Questions and discussion

- **0:10 – 0:15 What role does collaboration play in your projects?**
  - *What do you feel collaboration brings to your research?*
- **0:15 – 0:25 Can you describe the type of information that you exchange during your meetings?**
  - *Is the information that you share different for different phases of a meeting?*
- **0:25 – 0:40 How important is the sharing of computer data/documents in your meetings? Can you describe some examples?**
  - *When working with data/documents of that type, what information is important for you to communicate to your colleagues? How do you communicate that information?*
  - *When are data/documents important in your meetings?*
  - *What percentage of your meetings do you spend discussing data/documents?*
  - (leading question) Are gestures important when discussing artifacts
- **0:40 – 0:50 Break**
- **0:50 – 1:00 Can you give me some examples of how you share data/documents when you are in the same meeting room.**
  - *For that example, would you say that you were able to communicate the point you were trying to make? What worked, what didn't?*
  - How important is direct, physical interaction (physical pointing, SmartBoard)?
  - How important is multi-user interaction (not necessarily simultaneous, but rapid exchange of users who perform interaction)?
- **1:00 – 1:10 Can you give me some examples of how you share artifacts when part of the group is at a remote location.**
  - *How does your collaboration change when there are one or more remote participants?*
  - *For those who have been remote participants, do you feel the collaboration is effective? What works and what does not work?*
  - *For those who are local participants, do you feel that the collaboration is effective? What works and what does not work?*
  - What information is lost when interacting with artifacts at a distance?
- **1:10 – 1:20 Over the past five months, do you feel that your collaboration has changed?**

Wrap up

Well, we are coming up on the end of our 90 minute time slot. Does anyone have any thoughts or comments that they would like to add at this time on any of the topics we have discussed? (DISCUSS) Thank you all very much for your time and input. It has been an interesting discussion and you have provided me with a lot of information to consider. I will be using the information that you have provided me with today, combined with the information I gathered during my observations of your group, to help develop new tools that support collaboration more effectively. If anyone has any questions or would like to provide further input please let me know. Thanks again for your time.

15.4 Ethnography: Coding Scheme

Below is a list of the main primary and secondary codes for utterances and gestures used in our ethnographic study (Chapter 7). Note that this coding scheme was also used to analyze the video used in the gesture study described in Chapter 8.

Primary	Secondary	Description
UTTR	Utterance codes	
	OBJ	Deictic utterance about a physical object
	ART	Deictic utterance about a digital artifact
	FBK	Utterance that provides feedback (mmm hmmm, yup)
	ECL	Exclamation (wow, cool)
	QUE	Utterance that poses a question
	RES	Utterance that responds to a question
	STM	Utterance that makes a statement
	LAF	Laughter
GEST	Gesture codes	
	OBJ	Pointing at a physical object
	ART	Pointing at a digital artifact
	PER	Pointing at a person
	EMP	Gesture for emphasis
	ACK	Gesture to acknowledge
	LST	Gesture that makes a list (one, two, three)
	DES	Gesture that describes (draw with hands)
	ATN	Gesture to get attention
	ACT	Gesture that involves a computer action (click on a button)
	MAN	Gesture that manipulates or modifies an artifact

Table 16: Utterance and Gesture codes used in the study.

The primary and secondary codes are recorded, as well as the subject that performed the event and the time the event occurred. In addition, where a problem occurred (we know a gesture is not transmitted to the remote sites), the visual stream used to communicate gestures, and comments about the event are also made. In terms of visual streams, SMRT implies the gesture is made with the Smartboard, BODY implies a physical gesture with the arm, and COMP implies a gesture made with the computer mouse. Note that compound artifact events (see Section 7.1.3) can be communicated based on these individual events. For example, an explicit artifact gesture event occurs when all of the following occur:

- a deictic utterance event involving an artifact occurs;
- an artifact pointing gesture is made at that artifact;
- both events above occur at the same time; and
- both events above are made by the same subject.

Such an event can be seen in the coding scheme below. Note that the determination of the compound explicit artifact event can be (and is in our analysis) made mechanically. Thus if the low-level coding is performed consistently, high-level compound events are computed automatically. An example instance of an explicit artifact gesture event (made with the body) is highlighted in bold italics below. Note that the event is marked as a "problem" event as the coder noted that the gesture could not have been observed by the remote site.

Primary	Secondary	Subject	Time	Problem	Stream	Comment
UTTR	STM	S1	0:48:23			"…"
UTTR	STM	S2	0:48:25			"…"
UTTR	STM	S1	0:48:30			"…"
UTTR	STM	S3	0:48:35			"…"
GEST	MAR	S2	0:48:35		SMRT	scroll
ACTN	MAR	S2	0:48:35		SMRT	
UTTR	STM	S3	0:48:39			"…"
UTTR	STM	S1	0:48:42			"…"
GEST	MAR	S2	0:48:45		SMRT	scroll
ACTN	MAR	S2	0:48:45		SMRT	
UTTR	STM	S2	0:48:49			"…"
GEST	MAR	S2	0:48:53		SMRT	change page
ACTN	MAR	S2	0:48:53		SMRT	
UTTR	***ART***	***S2***	***0:48:59***			***"here is the equation"***
GEST	***ART***	***S2***	***0:48:59***		***BODY***	***point at new equation***
PROB	MIS	S2	0:48:59	MAJ		gesture missed
UTTR	STM	S2	0:48:59			"We have T sub i"
GEST	ART	S2	0:48:59		BODY	point at a term
PROB	MIS	S2	0:48:59	MAJ		gesture missed
UTTR	STM	S2	0:49:03			"minus little H sub i"
GEST	ART	S2	0:49:03		BODY	point at a term
PROB	MIS	S2	0:49:03	MOD		gesture missed
UTTR	STM	S2	0:49:09			"and the omega"
GEST	ART	S2	0:49:09		BODY	point at a term
PROB	MIS	S2	0:49:09	MOD		gesture missed
UTTR	ART	S2	0:49:16			"these are interesting…"
GEST	EMP	S2	0:49:16		BODY	general emphasis gesture
PROB	MIS	S2	0:49:16	MIN		gesture missed
UTTR	STM	S2	0:49:23			"…"
GEST	MAR	S2	0:49:51		SMRT	change page
ACTN	MAR	S2	0:49:51		SMRT	
UTTR	QUE	S1	0:49:51			"S4"
UTTR	RES	S4	0:49:53			"yes"
UTTR	STM	S1	0:49:56			"we can hear your,,,"

Table 17: Extraction from a coded meeting

15.5 Ethnography: Detailed meeting analysis

The ethnography presented in Chapter 7 provides an analysis of three of the eleven meetings observed in our study. These meetings were Meeting 3 (M3), Meeting 4 (M4), and Meeting 11 (M11). This appendix provides a more complete description of each of those meetings and the phases they went through.

15.5.1 Meeting 3 analysis

M3 was distributed with one participant at a remote site overseas (a hotel room). The meeting lasted one hour and fifteen minutes. The main topic of the meeting was the discussion of the data set and a mathematical model of the system that attempted to model the organization's processes. The model was instantiated as a computer simulation and produced numerical results. These results could be viewed in a number of ways, numerically and graphically.

From the perspective of the CoGScience Framework, there were two **sensory streams** used in this meeting. There was a moderate **fidelity aural stream**, utilizing an overseas phone connection to a hotel in Europe. There was a **high fidelity** (1024 x 768 pixels) application **visual stream** of the computer desktop (using VNC) sent to the remote collaborator. This allowed the collaborator to see any application running on the computer as well as any interactions that were performed using the mouse or the Smartboard. There was no visual stream that allowed the remote participant to see the other participants in the room.

The following paragraphs explore M3 in detail, with each paragraph denoting a specific phase of the meeting where there is a change in the way artifact interaction occurs. The start time of each phase of the meeting is noted at the start of the paragraph, with the interactions that took place during each phase described briefly in the paragraph body.

0:00:00 - The first phase of the meeting was technical setup. Although the research group is competent at using collaboration technologies, they struggled to connect to the remote user. One of the local participants (L1) was communicating with the remote participant (R) via cell phone to establish a connection. R used the hotel internet to obtain a network connection. After struggling with the technology for approximately 30 minutes, they eventually established a connection. Since this was an important meeting

and R was a key participant (the developer of the computational model), it was important that the meeting proceed. During this period there were no artifact focused communications.

0:32:12 - Once a connection was established, the meeting rapidly moved into a discussion phase. R connected to the local computer using VNC, and was able to see the local spreadsheet. The next seven minutes consisted of a discussion of the system that the model was trying to emulate. R had the most knowledge about the data set, and was explaining the data set in detail. R was directing L1 to manipulate the artifact (the spreadsheet) so that certain artifact features were on the screen. There were 53 utterances related directly to understanding the data, 16 of them referring directly to artifacts ("the bottom one", "the very top"). There were seven body language events related to observing the artifact, such as leaning forward toward the screen or getting up and moving closer to the screen.

0:39:24 - One of the local participants asked "…can you walk down the columns?" In response, L1 started pointing at artifacts and asking R questions about them. At this point we start to see composite artifact events (artifact utterances combined with artifact gestures) being generated. Three explicit artifact events were generated. R spent this time explaining the output of the computational simulation.

0:41:51 - R took control of the application and mouse and continued to explain the model output. This phase was very interactive, with local users asking questions and R answering, using the mouse as a gestural tool. During this period, there were 9 explicit artifact events ("down *here*") and 11 implicit artifact events. All of these events were generated using the mouse as the pointing device.

0:48:31 - L1 took control of the application and loaded a document that contained a number of simple visualization artifacts (2D graphs). R continued to describe his preliminary analysis of the model and the data, including exploration of the visualizations. L1 and R swapped control over the document on several occasions, at which time L1 changed the scale of the document so the participants could see the graphs better. During this phase, there were no explicit or implicit artifact events generated using the computer but there were two implicit artifact events generated physically (someone physically pointing at the screen). It is worth noting that R, the expert on the data set,

could not see any of these interactions and therefore had to work from the utterances only.

0:58:15 - The phase changed into a discussion about another graph. This phase resulted in a small amount of manipulation of the artifact (scrolling, changing pages etc.) but did not result in any gestural artifact interaction. Local users became confused as R referred to artifacts but did not point to them using the mouse.

1:06:50 – The original spreadsheet is reopened and a new tab is brought up with the simulation results of an initial model created by R. R describes the data produced by the simulation. At one point, L1 requested R to clarify an artifact utterance, at which time R began to use the mouse as a gestural interaction mechanism again. The discussion began to focus around understanding of the data set and model rather than a description of the model itself. This changed the interaction into a more dynamic and interactive phase, with R manipulating the artifact and using both explicit (4 times) and implicit (5 times) artifact events over a four and a half minute period. Much of the confusion was resolved through these artifact interactions.

1:12:57 - The artifact centric portion of the meeting was completed and the group started discussing next steps. There were no artifact or object related activities during this time. The meeting closed at 1:34:06.

Summary: By the end of the meeting, the research group appeared to have a basic understanding of the data set and the model R used to simulate the system. It appears that the research group gained some key insights about the system under investigation because of this meeting.

15.5.2 Meeting 4 analysis

M4 took place five days later and was a similar meeting in basic structure to M3. The goal of the meeting was to explore further the system being modeled and to validate the model that was being developed. One of the other participants had developed a second, independent mathematical model for the system, and this model was also explored in the meeting. The main difference between M4 and M3 in terms of meeting composition was that all participants in M4 were collocated. One additional member joined the group and the remote user from M3 was now on site (denoted as L2 in the following description). The phases of M4 are explored in more detail below:

0:00:00 - The meeting started with a description phase where L1 (the same L1 from M3) gave an update on the project status and discussed operational issues for the group. There was no artifact centric collaboration during this phase.

0:14:00 - L2 opened the spreadsheet containing the raw data and the original modeling results. There was a brief discussion about the data.

0:18:44 - L2 started to describe the data in some detail, with the goal of validating the model. Explicit and implicit artifact events were observed (14 times) and artifact manipulation (scrolling, changing worksheets, etc) occurred 8 times.

0:22:46 - L2 opened a text document that contained the raw output from the computational model. The output of the model was described, with some questions and comments from the other participants. Again, a significant number of artifact interactions were recorded (11 explicit, 14 implicit, 4 manipulate events). All of these interactions were generated on the computer using the mouse.

0:26:36 - L2 opened a document that contained the code for the computational model. This was used to explain how the model worked. This phase of the meeting also had a large number of artifact-centric events (11 explicit, 11 implicit, and 4 manipulate events). Although most of the 11 explicit gestures were made by L2 using the computer, three of them were physical gestures made at the plasma by another participant.

0:33:55 - The group's focus turned to a different approach to modeling. In order to explore this approach, the group switched to using the Smartboards. L1 started to write on the screen, outlining a set of mathematical equations for a new model. Actions included manipulating the application (creating new pages) as well editing the artifacts (the equations). This phase of the meeting had 2 explicit artifact events (both physical interactions), 9 implicit artifact events (all through the Smartboard), 7 artifact manipulations, and 11 artifact editing events.

Note that direct interactions with the Smartboard are coded as implicit artifact events. If a participant states "…v is the velocity" while writing it on the screen, this is captured as an implicit artifact event. The main reason for coding in this manner is that the writing gesture has much the same effect as underlining text while making the same utterance. It implicitly highlights the utterance.

0:37:56 - The group switched back to the laptop display, with L2 in control of the document. This was an active phase, as the group was dynamic and a number of people were asking questions and exploring the output of the initial model. During an 11 minute period there were a wide range of artifact interactions (21 explicit, 13 implicit, 14 manipulations, and 3 editing events). Three of the explicit artifact events were physical gestures with one of the gestures being used to identify a potential problem in the model.

0:48:41 - At this point, the interaction became very dynamic. The group switched back to the Smartboard and the second model, using it to determine a set of parameters that could be used to help determine if there was a problem in the first model. There were significant artifact interactions during this phase (5 explicit, 10 implicit, 2 manipulations, 9 edits).

0:50:53 – The group switched back to the laptop and the first model (3 explicit, 4 implicit, 4 manipulations), exploring the new parameter set.

0:56:12 – The group switched back to the Smartboard (6 explicit, 3 manipulations) to quickly explain the second model to one of the participants.

Between 0:48:41 and 0:57:00 13 of the 14 explicit artifact events were from physical gestures and were generated by more than one participant. Participants were very engaged in the meeting, as they appeared to be gaining an understanding of the problem with the model.

0:57:00 - The final artifact-centric phase of the meeting had the group exploring the first model (on the laptop) to identify the problem. Again, this was a very interactive and dynamic phase (22 explicit, 13 implicit, and 12 manipulation events). Seven of the explicit and 10 of the implicit artifact interactions were physical interactions where one or more of the participants were up at the Smartboard gesturing at artifact features.

1:05:22 - L2 identified the problem in the model and explained what the issue was and how it could be fixed.

1:13:51 - The artifact-centric portion of the meeting was finished and the group started discussing future project plans. At 1:15:00 the meeting finished.

Summary: The meeting appeared to be successful. The group had a clear understanding of the underlying data and was able to identify an important flaw in the

original model through comparison with a second model. Artifact interaction was used throughout the meeting.

15.5.3 Meeting 11 analysis

M11 was a very different meeting from M3 and M4 presented above. M11 focused on the discussion of two papers that were relevant to the group's research. The papers presented models that the group was considering integrating into their research. The papers were mathematical in nature, and much of the discussion revolved around the formulas and figures that were contained in the papers. The two papers were presented by two different participants, with both presenters at the local site. There were two remote participants.

From the perspective of the CoGScience Framework, this meeting was very similar to that provided in M3. There were two **sensory streams** used in the meeting. There was a moderate **fidelity aural stream**, utilizing Skype between the local site and the two remote sites. There was a **high fidelity** (1024 x 768 pixels) application **visual stream** of the computer desktop (using VNC) sent to the remote collaborator. This allowed the collaborator to see any application running on the computer as well as any interactions that were performed using the mouse or the Smartboard. There was no visual stream that allowed the remote participants to see the other participants in the room.

The meeting proceeded as follows:

0:00:00 – The first paper was presented by one of the local participants (L1). There was very little direct interaction with the computer or the Smartboard during L1's presentation, with the speaker going through a brief set of slides about the paper. There was significant discussion among the other participants through this phase, with more time spent discussing the paper than L1 presenting the paper. There were no artifact interactions during this phase.

0:12:40 – The group discussed the paper in a significant amount of detail. This phase consisted of discussion among all group participants but had no artifact interaction.

0:39:12 – L2 described the second paper, using the paper displayed on one of the plasma screens directly. L2 used the Smartboard extensively during this description, both to manipulate the artifact (scroll the document) and to mark up the document (underline something of interest using the digital pens). This phase of the meeting consisted of a set

of interleaved short description phases (1 – 8 minutes in length) followed by short discussion phases (1 – 5 minutes in length). During this phase there were 40 implicit and 25 explicit artifact interactions.

0:59:34 – The entire group discussed the paper. There were no artifact interaction events during this phase of the meeting.

1:15:00 – L2 finished describing the paper with a focus on some final terms in the model and how they were used. There were significant artifact interaction events during this phase (22 implicit and 9 explicit interactions).

1:20:25 – The group wraps up its discussion of the paper and discusses plans for next steps in the project. The meeting ends at 1:34:00.

Summary: Two papers were presented and discussed extensively. The research group's understanding of the two papers appeared to be strong, in fact strong enough to come to the conclusion that they would not be able to apply the approaches presented in the papers to their research.

15.6 Gesture Study: CoGScience analysis

This Appendix provides a more detailed application of the CoGScience Framework to the presentation utilized in the Gesture Study (Chapter 8 through Chapter 11). An overview of the collaboration task was given in Section 8.1. Here we provide a more detailed decomposition and discussion of the task domain, including the task characteristics of such a presentation.

15.6.1 The task domain

The main **task type** of a presentation is to deliver (**execute**) the presentation. That is, it is a **performance** driven task. That is, the goal is to **execute** the task (deliver the presentation) such that it convinces the audience that they should make a particular choice in the actions that they take. Although the main task is to execute (give the talk), the task also has aspects of a **choosing** task. That is, the speaker is trying to convince the audience that they should chose to take action. Since the task is not interactive in nature (it is a presentation, not a discussion) the task is more oriented towards **performance** than the group **choosing** the right outcome.

Considering the **choosing** task in more detail, from the presenter's point of view, there is a demonstrably "correct" choice and the goal of the talk is to convince the group that is watching the presentation to make the correct choice. This makes it an **intellective task** according to the CoGScience Framework. It is worth noting that from the audience's perspective, the "correct" choice may be unclear, which according to the CoGScience Framework would classify this as a **decision making** task.

When considering the **functions** required to accomplish this task, the CoGScience Framework suggests that two key functions are to **express ideas, engage** the audience, and **explain** a complex topic**.** In addition, the goal is to help the audience **decide** what action to take. In particular, we target this study at helping to increase our understanding of how gesture and facial expression affect the ability to **express ideas** and make **decisions**. In a normal distributed presentation, the ability to **discuss** would also be an important communication function to consider. In this study, we eliminate this function from consideration by controlling for it as part of the study.

Several **processes** are critical to this type of communication task. Processes that support the **conversation** include **engagement** and developing **trust**. Processes that support the **work object** include the ability to **create**, **modify**, and **manipulate** artifacts as part of the presentation. It is also necessary for the audience to be **aware** of the work space and to **monitor** how the speaker is interacting with that workspace. In fact, it is within this workspace that we intervene by controlling for how much workspace awareness subjects in different conditions have (see Section 8.3 for details). It is important to note that the workspace awareness required to perform this task is at a communicative level, not at an interactive or control level. That is, the workspace is not used to **coordinate** group work as it might be in tasks where the collaborators are synchronously working together on the artifacts in the workspace.

Finally, the CoGScience Framework also suggests that it is necessary to consider the human communication **channels** used to accomplish this task. As discussed in Section 2.2 and Section 2.3, it is the view of most researchers (across many domains, including computing science (CSCW, HCI), communications, linguistics, social psychology) that the gold-standard of group collaboration is face-to-face collaboration. The main reason for this relatively widely held view is that all of the human communication channels are

available in this form of communication. Not only that, but we have learned to use these channels effectively since birth. Thus, when one considers any communication scenario, it is not surprising that all human communication channels appear important. The benefit of the CoGScience Framework is that it provides a mechanism that allows us to consider these channels in the context of the functions and processes that need to be carried out during the communication.

From an aural perspective, the ability to **verbalize** is of critical importance. Also important to communicating ideas effectively and engaging the audience is the ability to be able to communicate **paralinguistic** information effectively (pitch, volume, intonation, rhythm, emphasis). Since artifact interaction is the focus of this study, all four forms of **gesture** (**manipulation**, **kinetic**, **spatial**, and **pointing**) are of high importance. Note that it is no coincidence that these channels map to the ArtifactManip (**manipulation**), EmphaticGesture (**kinetic**/**spatial**), ImplicitPointArtifact (**pointing**), and ExplicitPointArtifact (**pointing**) AOIs that we use in this study. One of the key experimental interventions of this study is of course gesture visibility, which we use as an independent variable. This study is designed to provide evidence about how important gestures are and in what contexts they help in the communication process. **Facial expression** is another human communication channel that is considered important in scientific presentations. It is also one of the independent variables that we manipulate in this study. **Body language** is considered less important, but is in some sense controlled in this study because only one condition provides body language (the YGYH condition). Since the presentation is not a one-on-one interaction, we do not consider **eye contact** as an important communication channel, although the speaker is always looking at the camera so some degree of **eye-contact** is present in this study when facial expression is visible. **Gaze awareness** (where an individual is looking) is relevant in this context as well, as the speaker looks at the artifacts as they are being pointed at and manipulated. This **gaze awareness** is missing in the No Head conditions. **Workspace awareness** is also highly relevant, and is provided to some degree in all conditions, with different degrees of interaction with the whiteboard workspace across the conditions. The richest **workspace awareness** environment is the YGYH condition, and our hypotheses predict that the YGNH condition will provide the next highest degree of **workspace awareness**

(through **pointing** at artifacts). The NGYH condition, through the movement of the head across the workspace combined with the **gaze awareness** provided by the presenter's face (the presenter looks at the artifacts that are being manipulated, even though the manipulation is not visible), also provides some **workspace awareness**. The NGNH condition provides a low level of **workspace awareness**, as subjects are aware of artifact **manipulations** made to the workspace (writing, circling) but do not have any of the other cues such as **pointing**, head position, or **gaze awareness**.

15.6.2 Task Characteristics

Using the **communication characteristics** of the CoGScience Framework, we further explore this collaboration task. The CoGScience task characteristics relative to this task are tabulated in **Table 18** and discussed below:

- Task Characteristics: The presenter is attempting to persuade the audience of the validity of certain scientific claims about global warming, and in so doing encouraging the audience to take action in their daily lives. The **material** or content of the talk presents abstract ideas but uses concrete representations of those ideas to communicate the concepts. The task is a **casual, informal**, high-level presentation to a general audience. A presentation is a **loosely coupled** task (presentations do not involve close interaction) and because it is primarily a narrative it is neither **creative** nor **exploratory** in nature. Although it is **difficult** to convince an audience of one's point of view, the collaboration among the participants is not **complex**. The **duration** of the presentation is relatively short (10 minutes).
- Environmental Characteristics: The presentation is in an academic environment and the **organizational norm** for such a presentation is to have the arguments questioned. There is little **competitiveness** during the presentation, although during a question and answer period this may not be the case. The topic of the presentation (global warming) is one that can result in **conflict**, and is viewed by some as both **emotional** and **urgent**.
- Group Characteristics: There is no **familiarity** between the participants and the presenter in this study and participants are isolated from the other members of the

group as a control. Since there is no group to interact with, **group familiarity**, **group size**, and **group skills** are not factors in this study. **Individuals** that participate in the study are senior undergraduate students, graduate students, research assistants, post-doctoral researchers, or faculty members.

	Characteristic	Categorization
Task		
	Material	Material contains abstract ideas, uses concrete mechanisms to explain those ideas
	Formality	Informal, casual presentation
	Coupling	Loosely coupled task, little interaction during presentation
	Exploratory	Not exploratory, chronicle of exploratory process
	Creativity	Not creative, narrative
	Difficulty	Convincing audience of presenters point of view is difficult
	Duration	Relatively short (10 minutes)
	Complexity	Not a complex task between collaborators
Environment		
	Org. norms	Academic - expect to be questioned
	Competitive	Not intrinsic in the communication, but often a side effect of presenting in an academic environment
	Urgency	Little urgency in communication, but inciting urgent action as a result of listening to the talk
	Conflict	Topic is considered controversial, so may cause conflict during the presentation
	Emotionality	An emotional topic that people feel passionate about
	Time of day	Participation occurred at all times during the work day
Group		
	Familiarity (Ind)	Participants did not know the presenter
	Familiarity (Grp)	Appeared as if participants were the only one's participating
	Size	Simulates a single remote participant in an office
	Composition	All participants have research skills, a mixture of seniority from senior undergraduates to senior faculty
	Individuals	All participants are senior undergraduates, graduate students, post-docs, research assistants, or faculty members

Table 18: Communication characteristics of a scientific presentation

15.6.3 Technology Domain

We use the CoGScience Framework to categorize and parameterize the technology domain of this study. There are two **sensory streams** communicated to the study participants, an **aural stream** and a **visual stream**. The **aural stream** is played back on standard desktop computer speakers. The audio on the video is a stereo audio signal, sampled at 44 kHz. It is encoded with an MPEG 3 codec as 192 kbps. As there is no transmission of the audio signal between remote sites (the presentation is pre-recorded),

there is no packet loss due to poor network connectivity. As there is no interaction between the subjects and the presenter, **feedback** (echo, often heard in poorly configured video conferencing sessions) and **delay** (due to network latency) are not an issue in this study. As a result, the **clarity** of the audio channel is high.

The **visual stream** is provided by a video of the presenter and in some instances, the workspace in which the presenter is working. The video is presented to the subjects on a 1024x768 pixel resolution LCD monitor. The video **fidelity** of the presentation is 640x480 pixels and is displayed in the centre of the screen with a black border around the outside. The video is recorded at 30 **frames per second** and was compressed using a moderate **quality** DIV-X MPEG-4 codec at a 1.5 Mbps bit rate. Like the audio, there is no **quality** loss or **delay** introduced due to network latency. Although the video was available as a high-quality MPEG4 video, the video codec used in this study was chosen intentionally to simulate the video **clarity** that is typical of H323 video conferencing (MPEG4 at 1.5 Mbps). The MPEG4 codec displays motion artifacts (blocky pixels, tearing of the image) when motion occurs rapidly in the video. The resolution was chosen to be higher resolution than standard definition H323 CIF video (352x288) because this resolution resulted in image **fidelity** that was considered too low to allow subjects to read text in the artifacts used by the presenter. We also wanted to avoid using HD quality video (1280x720 or 1920x1080 pixels). Although these resolutions are available in modern H323 video conferencing units, we wanted to simulate more traditional video conferencing technologies. In particular, it was desirable to maintain video **fidelity** levels that were similar to those utilized in the ethnography presented in Chapter 7. Thus, a resolution of 640x480 was chosen as a resolution that provided enough **fidelity** to read the text but was not of such a high **fidelity** that it was indistinguishable from content that would be provided on a laptop (e.g. 1280x720 pixels). The **field of view** covered by the video varied depending on the scene, from fairly tight framing of the presenter filling a significant portion of the screen to a more wide angle framing that included the presenter and the whiteboard used in the video. Given that the communication is one way, there is no reciprocity (by design).

15.7 Gesture Study: The NGYH condition

To provide visual information about facial expression without communicating information about the body language and gesture, we considered three possibilities. The first two methods were picture-in-picture (PIP) views of the presenter's head in the top left corner of the screen. These views were considered good candidates because they modelled a common communication technique used in traditional video conferencing systems. Two types of movement within the PIP window were considered, keeping a static PIP window in the top left corner of the screen and when the presenter moved out of the PIP window the presenter would not be visible. This simulates having a static camera on the presenter, with the presenter moving off camera on occasion. This has the benefit of simulating a standard H323 video conference environment, but does not provide facial expressions throughout the whiteboard scenes.

The second PIP method considered also used a PIP window in the top left of the screen, but kept the presenter's head centered within the PIP window as the presenter moved around in front of the whiteboard. This simulates having a camera follow the presenter around as he moves around in front of the whiteboard. This has similar benefits to the previous option, but also keeps facial expression on the screen at all times. This approach has the draw back of showing the original whiteboard behind the presenter as the presenter moves in front of the whiteboard, while at the same time has the digital version of the whiteboard on the non PIP section of the screen. Both of these options also have the important drawback of not having the same AOIs as the original video. Thus it is necessary to create a different set of AOIs, which in turn makes it much harder to compare AOIs across conditions.

The third method of displaying facial expression allowed the presenter's head to appear overlaid on the underlying video as the presenter moved around in front of the whiteboard. By creating a video mask that leaves very little space around the presenter's head, it is possible to create a "disembodied" head moving around the video as the presenter interacts with the whiteboard. This technique has the benefits of the presenter's head appearing in exactly the same location as where the presenter's head appears in the YGYH video, allowing us to use the same AOIs for analysis across all conditions. The

main disadvantage of this approach is the “unnatural” presentation of a disembodied head as part of the presentation.

All videos utilized the NGNH video as their basis, using the video editing software Final Cut Pro to overlay the picture-in-picture box and the “floating head” videos over the NGNH video. The videos were edited by a student from the Film School at Simon Fraser University, and therefore the quality of the mask in the floating head video was of very high quality.

Although it seems intuitive that the PIP approaches would provide a good experience in communicating facial expression, the movement of the presenter caused serious problems for both of these methods. Since the first method simulated a static camera, the presenter spends a significant amount of time outside of the PIP window. While the second PIP technique kept the presenter within the PIP window, having the diagram on the whiteboard both behind the presenter as well as next to the presenter was very disconcerting. In addition, both PIP videos present the facial expression in a completely different location than in the YGYH video. This raises concerns about the impact of keeping the facial features focused in the top left part of the screen rather than closer to the artifacts that are manipulated. That is, it would be difficult to tell whether or not it was the location or the visibility of the facial features that cause differences in subject performance. In addition, having different AOIs across conditions makes analysis significantly more difficult.

The floating head video does not suffer from any of these problems. The masked presenter head is displayed in exactly the same location as it is in the other condition and the same AOIs can be used for analysis. This allows the most direct comparison and of the three possible videos, controls for the most extraneous variables (different head positions, potentially confounding duplicate artifact information being presented). For these reasons we chose the floating head video. An image from this video can be scene in Figure 41 (p. 204).

15.8 Gesture Study: Post-study questionnaire discussion

A post-study questionnaire was give to all subjects at the end of the second video. This questionnaire posed questions to the subjects about their understanding of the contents of the presentation as well as the structure and information contained in the artifacts used

during the presentation (as described in Section 8.6.1.2). The questionnaire consisted of nine questions. The first eight questions were about how artifacts were used in Act 2 and Act 4 of the presentation and are discussed in more detail below. The ninth question was an open-ended question that allowed the subjects to provide general comments about the presentation.

Act 2 and Act 4 both contain extensive presenter interaction with an artifact on the whiteboard, in this case the Pascal's Wager diagram shown in Figure 50 (p. 219). Question 1 poses an open-ended question that enables us to analyze qualitatively the subject's understanding of the use of the diagram and in particular, whether there is understanding that the diagram's primary role is to make help make a decision when faced with uncertainty. It also allows us to determine whether a subject has been exposed to Pascal's Wager or a similar decision making tool in the past.

Question 2 and Question 3 deal with the **structural** components of Pascal's Wager diagram and the role they play in the presentation, and in particular what the roles of some of the diagram components play in the presentation. Question 2 asks *"What do the rows in the diagram represent?"* and Question 3 asks *"What do the columns in the diagram represent?"* Although the questions are open-ended, we chose this over providing a multiple choice answer in order to avoid giving subjects hints about the correct answer. The correct answers to the questions were determined by the researcher, confirmed by the author of the video (the author of the video was sent the questionnaire and provided answers), and further validated by the questionnaire inter-coder reliability process described in Appendix 15.9.3. Both questions were scored as a total out of two, where two of the two marks were given for a correct answer and one out of two was given for a partially correct answer.

Question 4 and Question 5 deal with the **information** aspects of the diagram. Question 4 asks subjects to describe one of the key informational aspects about the presentation – "*According to the presenter, what are the potential risks of taking action against global warming*"? In the pre-study trial and the early parts of the study itself, it was noted that this question was too vague (see Section 9 for a discussion of this change). The question was later refined to contain a second phrase – "*Which of these risks did the presenter list in the four quadrant diagram*?" We use this question because the answer (Economic

Harm) is an artifact in the diagram (the top left quadrant in Figure 50), is created dynamically during the presentation (through writing), and is referred to repeatedly by gestures. Due to our experimental interventions, subjects from different conditions will see this information communicated in different ways. In this manner, we can measure the differences in understanding about the informational aspects of the diagram across conditions. It should be noted that Question 4b is in some sense a trick question, as we ask *"… what risks …"* where it is clear from the diagram in Figure 50 that only one risk is listed (Economic Harm).

Question 5 is similar to Question 4 except it measures understanding of a topic that is communicated by a different area of the diagram. Question 5 reads, "*According to the presenter, what are the potential risks of not taking action against global warming? Which of these risks did the presenter list in the four quadrant diagram?*" Question 5 is also a two-part question, with the first part asking for a general answer and the second part asking for specific information from the diagram. Like Question 4, the second part of the question was added part way through the study. The information required to answer the question can be found on the diagram, is created dynamically during the presentation, and is referred to several times with pointing gestures. Like Questions 2 through 4, Question 5 has a clear, correct answer (environmental, political, social, public health, environmental disasters) and can therefore be measured effectively. Both parts of Question 4 and the first part of Question 5 were scored out of two marks. Part two of Question 5 was scored out of five marks, as there were five risks listed in the diagram.

Questions 6 and 7 measure a subject's understanding of the **argument** that the presenter is making. The logical argument posed by the presenter is that the evidence that humans are causing global warming is very strong, and therefore the bottom row is much more likely. In addition, the presenter attempts to demonstrate the risks of not taking action are much worse than the risks of taking action. This argument is presented in Act 6 (see Figure 67), as the presenter both presents his scientific argument to the audience as well as manipulates the artifact/diagram by erasing and redrawing the line that represents the likelihood of whether humans are causing global warming (the line that divides the true row from the false row is moved upwards). The questions asked are "*According to the presenter, which of the rows in the diagram is most likely to occur? What rationale*

does the presenter use to justify this position?" and "*According to the presenter, which of the columns in the diagram has the most significant risk? What rationale does the presenter use to justify this position?*" Both are two-part questions, with the first part of both questions enabling the measurement of the understanding of the concept and the second part of the question enabling the measurement of the rationale the presenter uses to justify his position. Both questions have concise answers, although they are not as explicit as Questions 4 and 5. The answers cannot be seen directly in the diagram, and it is therefore the argument that must be understood to be able to answer these questions correctly. Both parts of Question 6 and Question 7 were scored such that a subject could score at most two marks on each part (eight marks across both questions). Like Question 4 and Question 5, the second part of these questions was added part way through the study. Like all questionnaire questions, the correct answers have been validated by the inter-coder reliability process presented in Appendix 15.9.3.

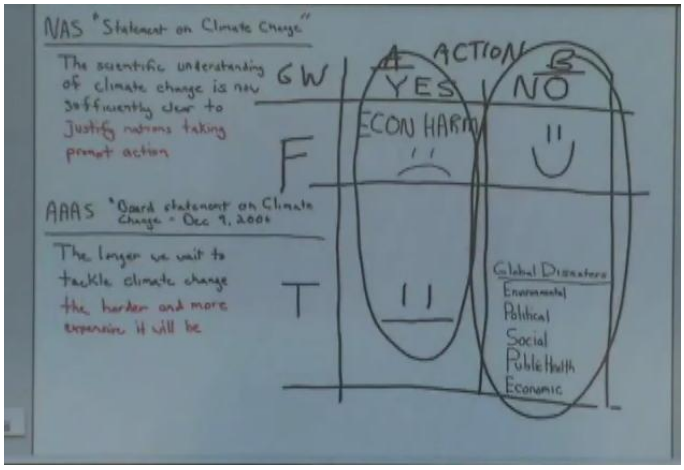


Figure 67: Act 6 Video after manipulations

Question 8, like Question 1, is an open-ended question that allows us to qualitatively assess the overall understanding of the argument made by the presenter. "*According to the presenter, what is the key unknown in trying to understand what action we should take to deal with global warming?*" This question is designed to measure the subject's understanding of where the uncertainty lies in whether we should take action against global warming or not. This question is scored out of two marks.

The complete questionnaire is provided Appendix 15.13.3.

15.9 Gesture Study: Inter-Coder Reliability

15.9.1 Scene Inter-Coder Reliability

We tested the validity of our scene decomposition and AOI assignment through an inter-coder reliability analysis. Inter-coder reliability of how coders divided a whiteboard-based act into scenes was determined by asking three different coders to divide Act 2 of the video into scenes. The first coder was the experimenter, who used a highly accurate video editing tool to determine a start and end time for each scene (start time recorded in milliseconds). We then compare the start times chosen by two other coders (other HCI researchers in the Chisel research lab), who used a normal video player (accurate to the second). Act 2 was chosen because it was representative of the whiteboard sections, with both artifact gesture and dialogue. At the same time, it was the most structured of the three whiteboard sessions, making it attractive for testing coder reliability.

In preparation for the study, an initial analysis of Act 2 was performed, dividing it into thirteen scenes. Eight of these were gestural pointing scenes, two involved no pointing (the presenter standing and talking), and three were post-action scenes. All of the scenes in Act 2, with the exception of the "post-action" scenes, are between one and three seconds in length. As discussed in Section 8.3.1.3, a post-action AOI is used to capture gaze information about artifacts in scenes where the action is no longer referring to the artifact. In some instances, there are sections of the video that are transitions between actions. We call these post-action scenes, as even though they do not capture any specific action we still need to capture gaze data about what the subject is looking at (often involving an action from the previous scene). Since we do not want to compromise how a scene captures a single action, it is necessary to create scenes that capture the transition periods between actions (a post-action scene). Post-action scenes are typically very short, with the three post-action scenes in Act 2 being 699, 874, and 626 milliseconds long.

Two other video coders were asked to consider Act 2 and determine the beginning and end of each scene. The coders were told that each scene should contain a single action made by the presenter, that there should be no gaps between scenes, and that there should be between 5 and 25 scenes. The coders were provided with a worksheet to enter the start

time and end time of each scene. They used a video playback tool with a time granularity of one second.

Although we are looking for agreement among coders, it is not reasonable to expect exact agreement for the start time for a given scene. This is especially true given that the granularity of the start time used for the study scenes is accurate to 10s or 100s of milliseconds while the granularity of the video playback tool used by the coders is one second. On seven of the eight gestural scenes in Act 2, there is very high coder agreement. Both coders were within 482 milliseconds of the experimental scene start time on average, with a minimum difference of 142ms and a maximum difference of 858ms. Note that such differences can be partially attributed to the accuracy of the playback software alone. For example, the study start time for Scene 2-11 is at 1:32:858 in the video (according to the experimenter's coding). Coder 1 chose 1:32 as the start time for this scene and Coder 2 chose 1:33. This subtle difference results in a timing difference of 858 milliseconds for Coder 1 and 142 milliseconds for Coder 2. Note that this single scene is responsible for both the largest and smallest coding difference.

On the one gesture scene where there was no agreement, one of the coders did not consider the action a separate gesture while the other coder did. This gesture was a subtle one, with the presenter moving his hand only slightly lower to point at a different artifact. The coder who did code this as a separate gesture agreed with the experimental scene coding time within 162ms. Based on this analysis, it appears that the coding process used for dividing scenes into gestural actions is robust.

Neither of the coders captured the three post-action scenes, as the transition times between scenes were included in one of the surrounding scenes. This is not surprising, as we did not consider post-action scenes necessary until we had performed some pre-study trials and realized and the importance of capturing information about post-action AOIs across multiple scenes. Of the two scenes that did not contain gestural actions, one coder captured both of them (with differences of 12 and 654 milliseconds) while the other coder captured one of them (654 millisecond difference). The coder that did not capture the non-pointing scene considered the non-pointing part of the action as part of the previous pointing action. This coder did capture the next pointing action accurately, implying that the coder missed only this one scene.

Overall, it appears that the scene coding process is robust. Of the ten gestural and non-gestural scenes (excluding the post-action scenes), both coders coded nine out of the ten scenes in the same way as they were coded for the experiment. For all of the scenes that were coded, the timing difference between the two external coders and the experimental scene coding was small and well within the difference one would expect through inaccuracy in the video playback tools used. Each coder failed to code one of the scenes correctly, but each coder missed a different scene and when a coder did miss a scene, the other coder coded it accurately. Thus, even for the scenes that were missed by one of the coders, the experimental coder and the other test coder coded them in the same way. The instrument used to in the scene inter-coder reliability protocol is given in Appendix 15.10

15.9.2 AOI Inter-Coder Reliability

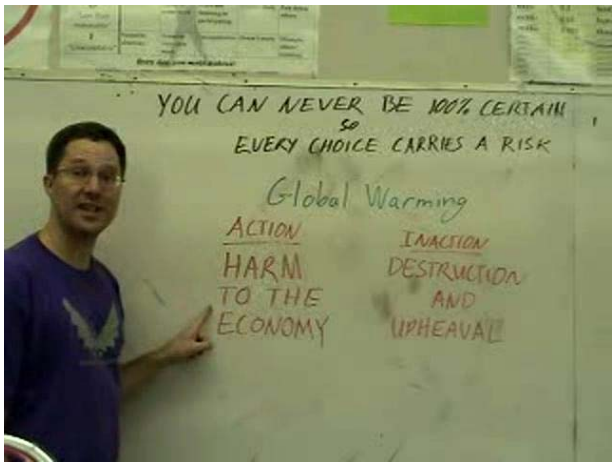


Figure 68: An example scene used for AOI inter-coder reliability

We tested the validity of our AOI assignment with an inter-coder reliability analysis. The same three coders who performed the scene inter-coder reliability were also presented with six scenes from the same section of video they had just coded. A scene was presented by providing an image from the video, the start time of the scene, and the phrase that was spoken at the time the frame of the video was captures. An example image from one of the scenes can be seen in Figure 68. The phrase associated with this scene was "… risking the possible harm to the economy that the sceptics warn us about …" The entire inter-coder reliability protocol is given in Appendix 15.10.

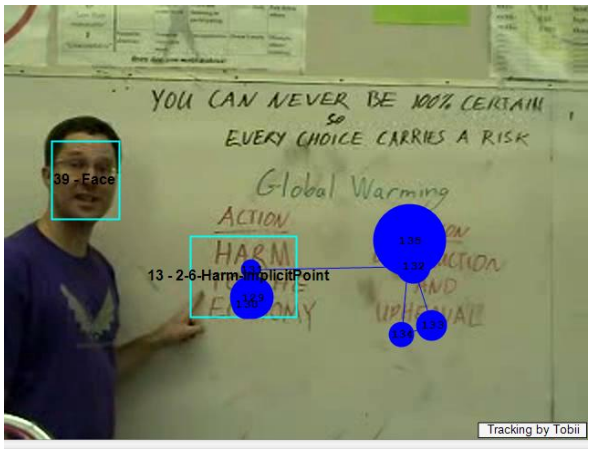


Figure 69: AOIs for Scene 2-6, as created by the experimenter and used in the study

As in the scene analysis, the experimenter created an AOI baseline. The other two coders were asked to create a set of AOIs that they felt the observer watching the video might look at during the scene. They were asked to draw a box around any area of the image that they felt the observer would focus a significant amount of attention. For all six scenes, coders were asked to create scenes that had at least one AOI and no more than eight AOIs. They were also told that only a small percentage (10 – 20%) of the image should be covered by AOIs (we wanted to create small, precise AOIs). Scenes with both gestural interaction and scenes with just dialogue were coded.

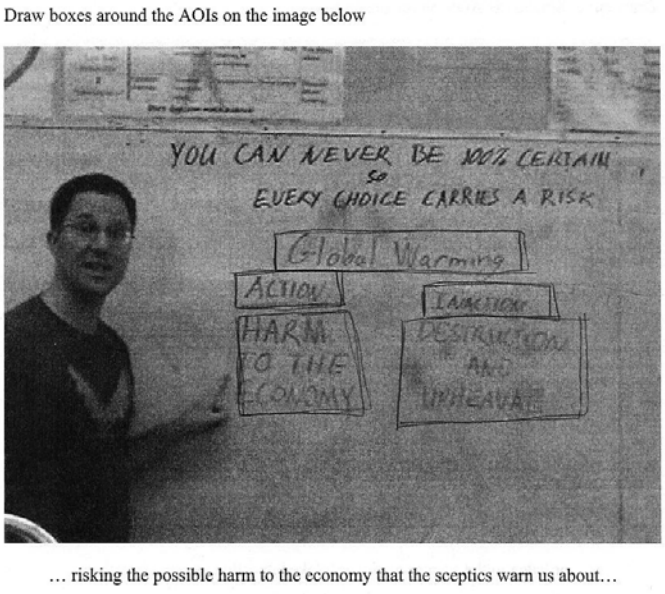


Figure 70: AOIs drawn by Coder 1 for Scene 2-6

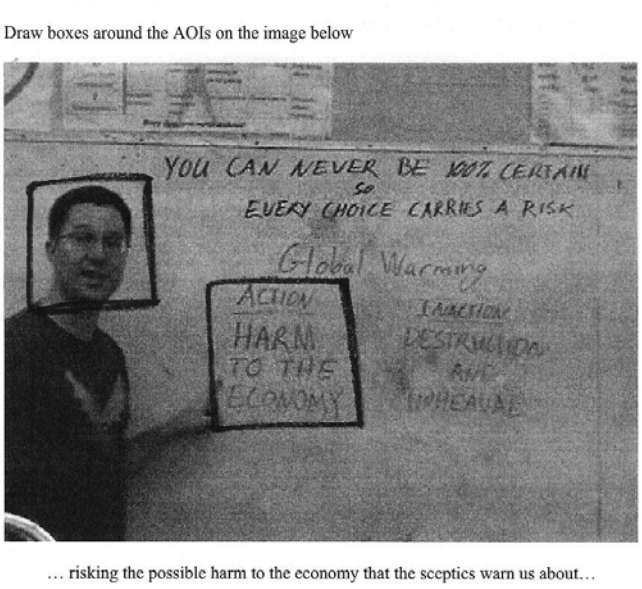


Figure 71: AOIs drawn by Coder 2 for Scene 2-6

Figure 68 shows the AOIs created by the experimenter and as used in the study for Scene 2-6. This figure also shows eye fixation data for one of the subjects in this scene (eye fixation data and eye tracking is discussed in more detail below). Figure 70 and Figure 71 show the AOI regions as created by Coder 1 and Coder 2 respectively. Note that both coders created AOIs around the "Harm to the Economy" artifact on the screen, with the only major difference being the size of AOI and how tightly it bounded the artifact. In general, Coder 2 had a much looser bounding criterion for the AOI. In addition, Coder 1 defined AOIs not only around the artifact being referred to by the gesture, but also around other artifacts in the scene (other text phrases in the diagram). Also note that Coder 2 created an AOI around the speakers face while Coder 1 did not, although this is not always the case. In fact, in all scenes both coders created one FacialFeature AOI and at least one artifact AOI (with the exception of no FacialFeature in Scene 2-6). A summary of the number of AOIs and their types created by Coder 1 and Coder 2 are given in Table 19.

Scene	Coder 1		Coder 2	
	FacialFeature	Artifact	FacialFeature	Artifact
2-1	1	2	1	1
2-4	1	4	1	1
2-6	0	5	1	1
2-8	1	2	1	1
2-12	1	2	1	1
2-13	1	2	1	1

Table 19: Number of AOI types for Coder 1 and Coder 2 for each scene tested

Since our primary concern is acquiring accurate fixation data about the artifacts involved in the interactions during the presentation, the original AOIs created for the experiment were closely bounded on the artifacts. This is very similar to the AOI encoding provided by Coder 1. This does not mean that we disregard the AOI encoding developed by Coder 2. In fact, in many of the cases where Coder 2 created a single, large AOI, Coder 1 had multiple AOIs that covered the same region. Since Coder 1's AOI provides us with more accuracy in terms of which artifacts are being fixated upon, we use the smaller, more accurate AOIs in this study.

Another important factor to consider is whether or not the study should contain AOIs around the many artifacts that exist in the scene. Since we are primarily concerned with fixations that results from the visual actions of interest (gesture and facial expression), this study only utilizes AOIs that are relevant to facial expression (or lack of it) and gesture. We do not make use of AOIs around artifacts that are not the immediate (or near term) referent of a gesture. For example, in Scene 2-6 (Figure 69) we do not have AOIs around any of the artifacts except those that are being referred to by the gesture being made during that scene. The one exception to this is of course the "post-action" AOIs discussed in Section 8.3.1.3.

In summary, the AOI coding scheme utilized in this study was initially based on our pre-study trials. We presented two independent coders with several scenes from the video, and analyzed their AOI selection. Although they differed slightly, they differed primarily in how tightly bounded the AOIs were on the artifacts in question. Thus, our AOI selection appears to be consistent across the coders tested. Since we are concerned with direct fixations as a result of gestural interaction, we utilize AOIs that are tightly bounded on the artifacts in question.

15.9.3 Questionnaire Inter-Coder Reliability

We tested the validity of the questionnaire answers in much the same way as we tested the Scene (Appendix 15.9.1) and AOI (Appendix 15.9.2) inter-coder reliability. Given that most of the answers to the questionnaire were completely available in the information that was presented in the video, finding the correct answers is straightforward if the video is referred to carefully. The same video coders used for Scene and AOI inter-coder reliability testing were asked to answer the questionnaire, using the video as much as required. The questionnaire was also given to several test subjects in pre-study testing. Lastly, the original author of the video was also asked to provide answers to the questions.

In almost all cases, the questions were answered the same, with some minor discrepancies. Two of the questions in particular were problematic. The question about what the rows in the Pascal's Wager diagram represented were not correctly answered by all coders despite being able to refer to the video as often as required. The questionnaire asks "*What do the rows in the diagram represent?"* The correct answer to this question is quite subtle in that the correct answer is *"Whether or not humans are the cause of global warming (T or F)".* Several of the coders interpreted the answer as *"Whether or not global warming is really occurring (T or F)"*. Despite this discrepancy between coders, this question was not changed for the study as it was thought that the experimental intervention might result in a significant difference in getting this question partially (global warming occurring or not) or completely correct (global warming caused by humans or not).

The second problem question was the last question on the post-study questionnaire. This question was intended as an open ended and interpretive question that was targeted at getting the subject to think of the "big picture" of the presentation. That is, there was no explicit answer given in the video and each subject was required to interpret the overall point the presenter was trying to make. The question reads *"According to the presenter, what is the key unknown in trying to understand what action we should take to deal with global warming?"* The presenter suggests that the key unknown is whether or not humans cause global warming, and his position is that the evidence is strong that human's are indeed the main cause. He demonstrates this in Act 6 with a key artifact

interaction (erasing and redrawing the line), moving the line between T and F in the diagram up higher. This question is designed to explore whether this key artifact interaction changes the interpretation across the conditions. Since there is no explicit correct answer for this question, it is not surprising that the answers across our coders are slightly different. Since we were looking for primarily a qualitative difference in this question, this question was also left unchanged.

15.10 Gesture Study: Scene and AOI inter-coder reliability protocol

In this section we provide the document given to the coders for both the scene and AOI inter-coder reliability tests. Coders were asked to read the document. An assistant was available to guide the coders through the process. The assistant only provided details on how to carry out the process of filling out the coder questionnaire and did not provide further details on what to code (that was left to the description in the document alone).

Overview:

We are performing a study in which we are trying to determine what subjects look at when watching someone make a presentation that involves a physical artifact (in this case a diagram on a whiteboard). We are using a video of a presenter giving a talk about global warming for this study.

We are interested in determining the following:

1. If subjects watch gestural actions of the presenter (movement of the hand).
2. If subjects watch facial features of the presenter.
3. What other parts of the video observers might spend a significant amount of time watching.

Your Task:

Your task is to divide the video clip up into analysis components such that analysis can be performed on the video in a consistent and methodical manner across many subjects. In order to perform this task, we would like you to perform the following two steps. In Step 1, we ask you to divide a video clip that is approximately 30 seconds long into a set of "scenes" where each scene consists of the presenter performing a single action (the action of pointing at an object, the action of standing and talking, etc.). In Step 2, we present you with a set of images from example scenes and ask you to determine Areas of Interest (AOI) where you think an observer of the video would focus their attention.

These two steps are described in more detail on the following pages.

Step 1: Scene Creation

You have been provided with a video clip that is approximately 10 minutes long. We would like you to divide a 30 second subset of this video clip into scenes as described below. The portion of the video we would like you to consider starts at the 1:10 (one minute and 10 seconds) mark of the video and ends at 1:35 (one minute and 35 seconds). The person helping you with the coding will make sure the video is in the correct location and will assist you if you need to start, stop, or rewind the video.

Your task is to divide the video clip up into contiguous scenes (over time) where each scene consists of a single logical action made by the presenter (e.g. the presenter points at something, the presenter is standing still speaking, the presenter is using his hands for emphasis).

1. Each scene should represent a single action made by the presenter (talking, pointing, gesturing).
2. Each scene should have a start time and an end time.
3. The start time of one scene should be the same as the end time of the previous scene (the scenes should be contiguous). Scenes can contain no action if there are pauses in the presentation that do not contain interesting actions.
4. The video clip should contain at least five and no more than twenty-five individual scenes.

Please write down the start and end time of each scene. Use your own judgement as to what you feel is a single action by the speaker. There is no correct number of scenes other than we are not interested in a scene granularity that would result in more than twenty-five scenes. You should feel free to start, stop, rewind, and re-watch the section of video as many times as you feel are necessary in order to determine appropriate scenes.

Scene Creation

	Start Time	End Time
Scene 1:	_____	_____ .
Scene 2:	_____	_____ .
Scene 3:	_____	_____ .
Scene 4:	_____	_____ .
Scene 5:	_____	_____ .
Scene 6:	_____	_____ .
Scene 7:	_____	_____ .
Scene 8:	_____	_____ .
Scene 9:	_____	_____ .
Scene 10:	_____	_____ .
Scene 11:	_____	_____ .
Scene 12:	_____	_____ .
Scene 13:	_____	_____ .
Scene 14:	_____	_____ .
Scene 15:	_____	_____ .
Scene 16:	_____	_____ .
Scene 17:	_____	_____ .
Scene 18:	_____	_____ .
Scene 19:	_____	_____ .
Scene 20:	_____	_____ .
Scene 21:	_____	_____ .
Scene 22:	_____	_____ .
Scene 23:	_____	_____ .
Scene 24:	_____	_____ .
Scene 25:	_____	_____ .

Part 2: Determining Areas of Interest (AOI)

In Part 2 we ask you to determine Areas of Interest (AOI) for several scenes. On the following pages you have been given a set of six images that represent six different scenes in the video you have just watched. For each scene, identify AOIs within the scene that you feel the observer watching the video will look at during that scene. An AOI consists of a **box** that encloses an area of the image where you feel that the observer might focus their attention during the scene. Recall that in this study we are primarily interested in determining what the observer is watching on the screen while the video is playing. We ask that you define AOIs for each image following the guidelines below. For each scene, we provide a portion of the presenter's dialog for context. If you want to refer back to the video to get more information about the scene feel free to do that. At the top of each image the time within the video is noted so you can easily find the scene being coded.

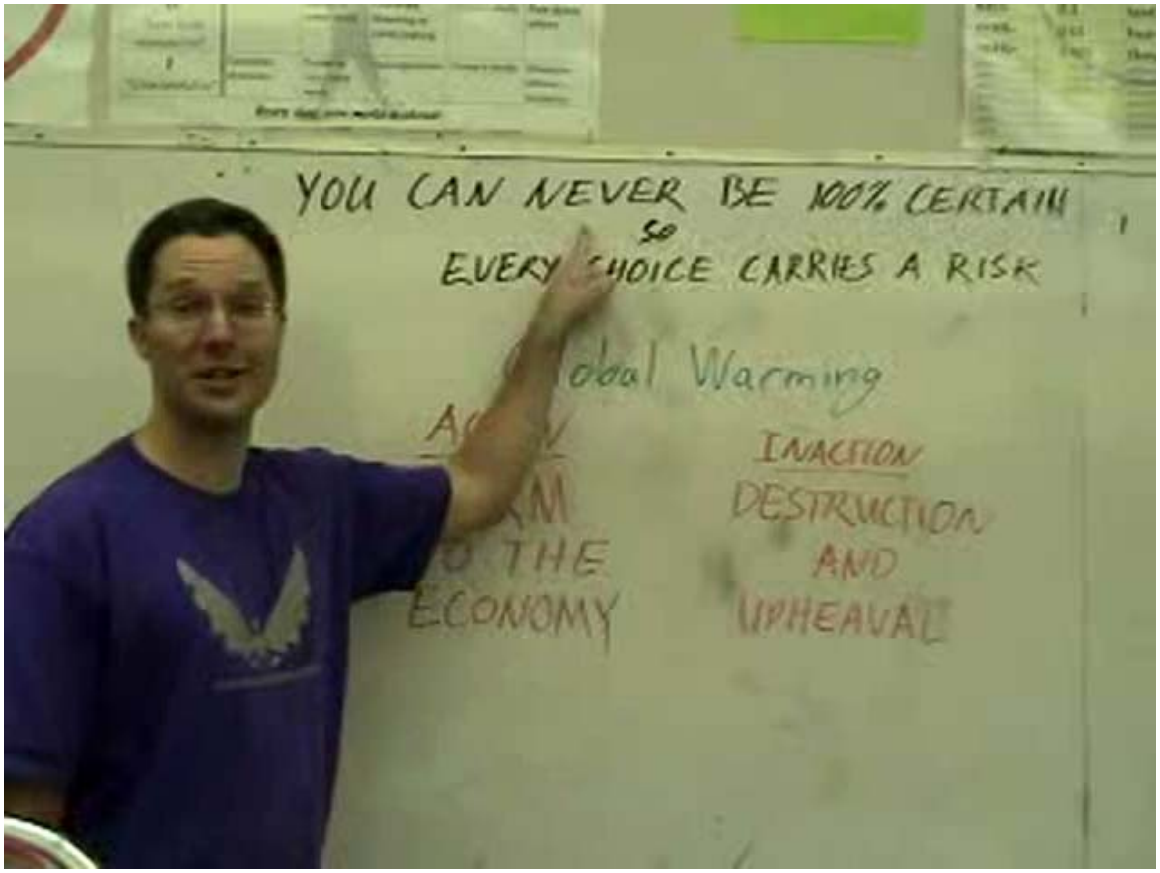
Guidelines:

1. Draw a box around any area of the image that you feel the observer would focus a significant amount of attention during that scene.
2. Each scene should include at least one AOI and no more than eight AOIs.
3. Only a small percentage of the each image should be covered by AOIs (10% - 20%). For example, don't use AOIs that cover the entire whiteboard, but instead use one or more AOIs to cover a specific area of the whiteboard that you think the observer will look at during a scene.

The images (and the presenter dialog for context) are given on the following pages. Please draw boxes around the AOIs on the images of each of the six example scenes.

Scene 1: (time = 1:11)

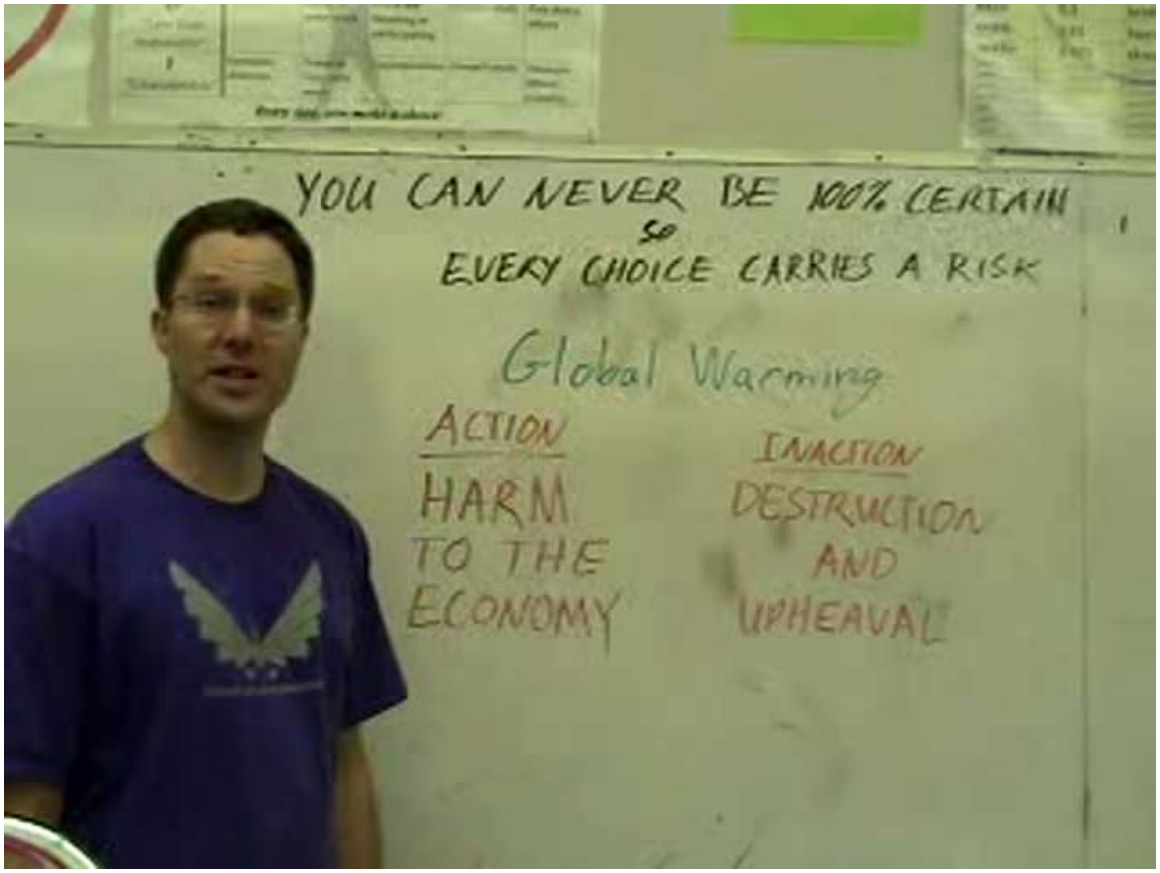
Draw boxes around the AOIs on the image below



First off, no one is perfect, so every choice you make brings with it…

Scene 2: (time = 1:16)

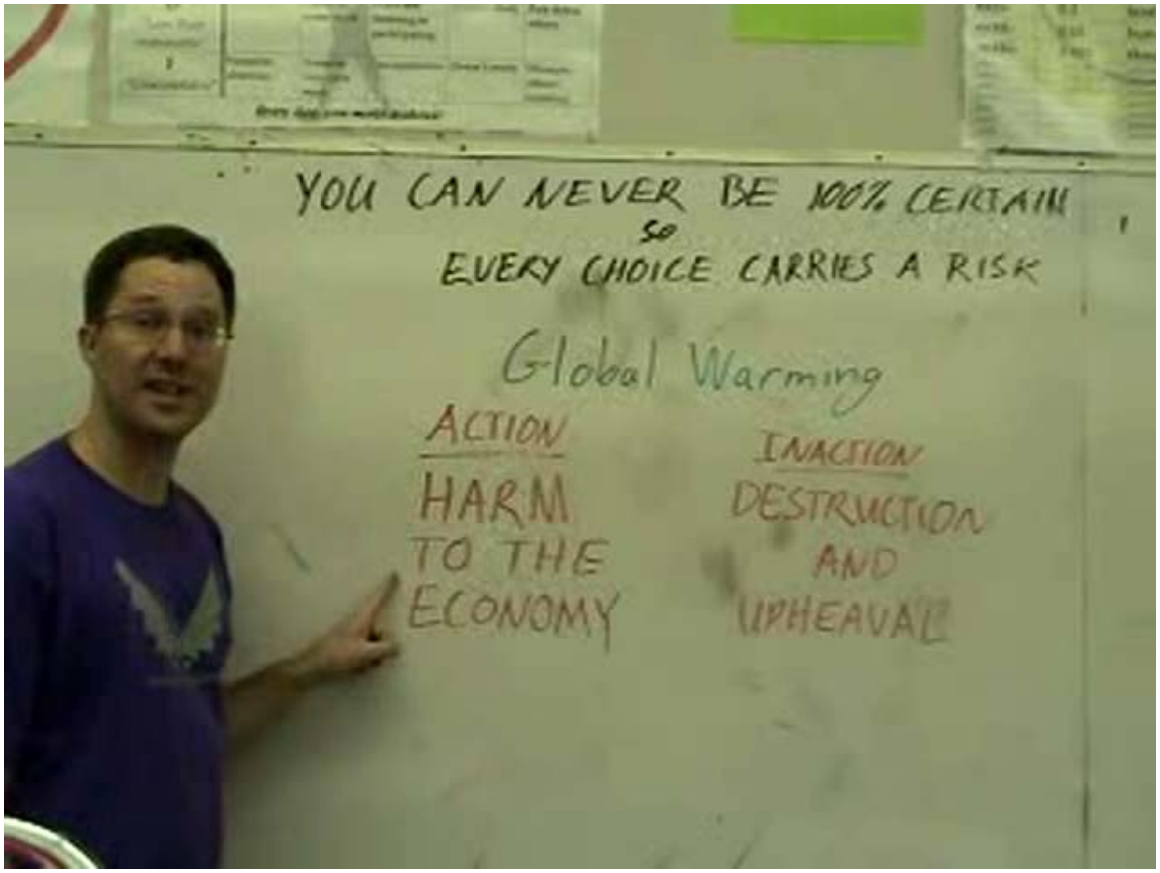
Draw boxes around the AOIs on the image below



… given that, which risk would you rather take, …

Scene 3: (time = 1:21)

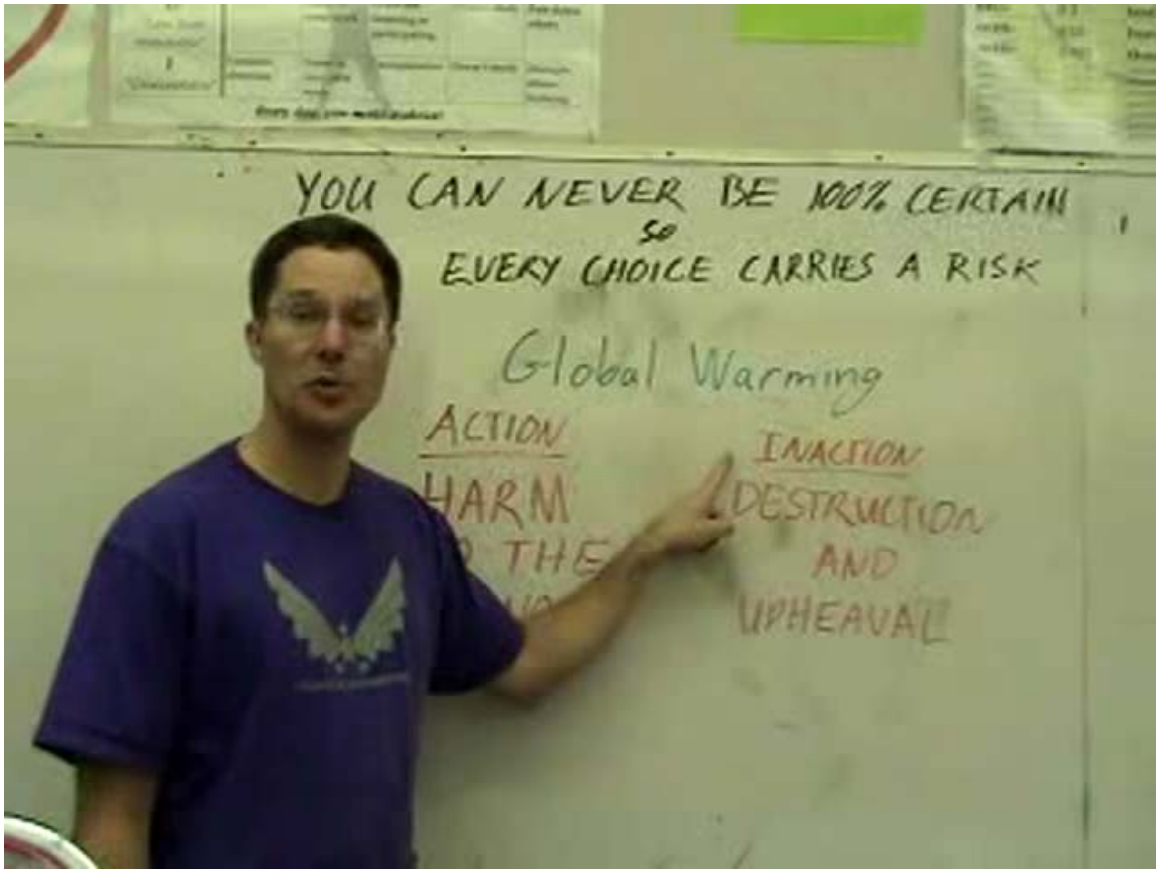
Draw boxes around the AOIs on the image below



… risking the possible harm to the economy that the sceptics warn us about…

Scene 4: (time = 1:25)

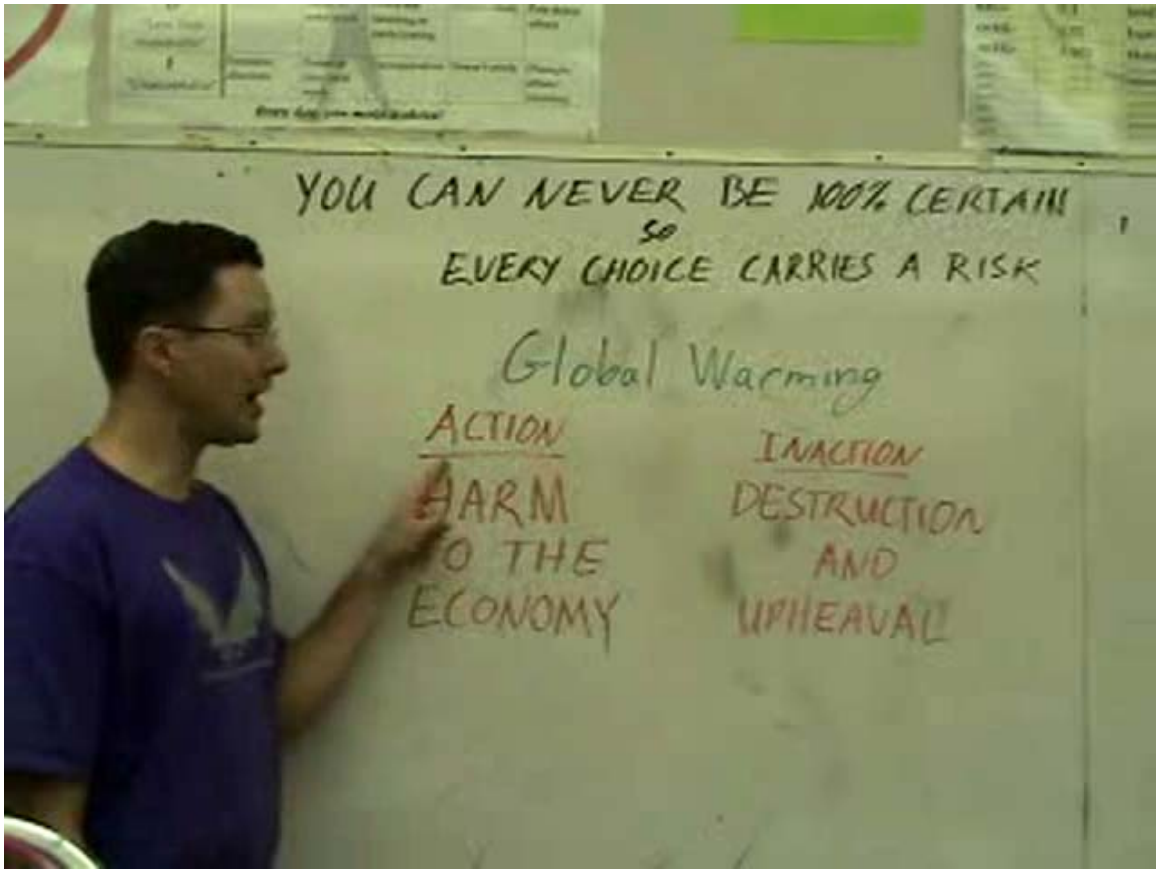
Draw boxes around the AOIs on the image below



… or listen to the sceptics, and don't take big action now, risking the …

Scene 5: (time = 1:33)

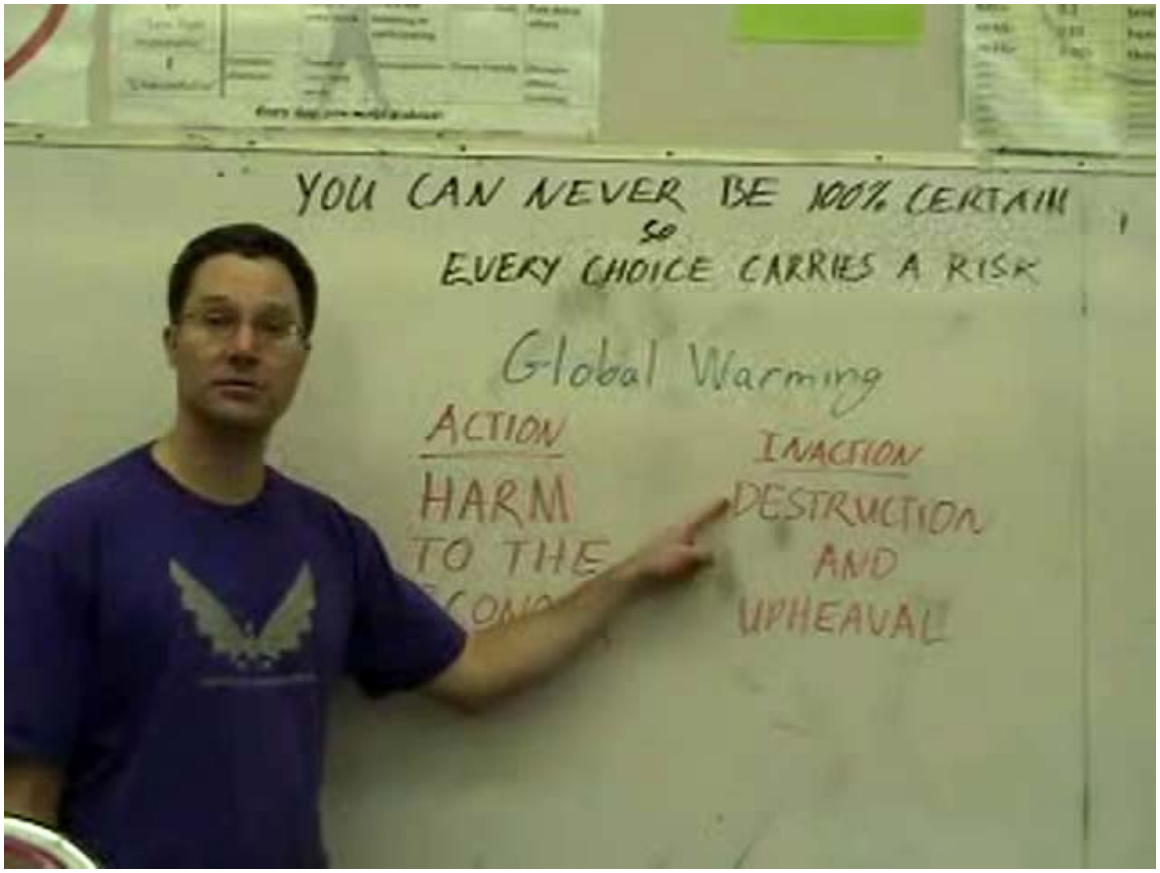
Draw boxes around the AOIs on the image below



… the risk of taking action, or …

Scene 6: (time = 1:35)

Draw boxes around the AOIs on the image below



… or the risk of not taking action.

15.11 Gesture Study: Recruitment letter

Subject: Communicating as a scientist – a research study

Are you a RESEARCHER that can LISTEN as well as you TALK?

Participate in a study of how researchers communicate and show us it is true!

As part of a research project in the Department of Computer Science, we are seeking people to participate in a research study to help us understand how researchers communicate. We are searching for volunteers who are active members of a research project at the University of Victoria and are either a:

- Senior undergraduate student (4th year)
- Graduate student
- Research assistant

- o Post doctoral researcher
- o Faculty member

Participation will require only 45 minutes of your time, will take place on the University of Victoria campus, and can be scheduled at your convenience. Participants in the study will be given two movie passes as remuneration for their time.

If you are interested in participating in this study, please contact Brian at bdcorrie@csc.uvic.ca.

15.12 Gesture Study: Observer notes page

Date:
Subject Name:
Start Tape Location:
Subject Arrival Time:
Camera Start Time:
Camera Stop Time:
Stop Tape Location:

Document Check List:

- Consent form:
- Pre-study questionnaire
- Mid-study questionnaire
- Post-study questionnaire
- Comments document

Comments:

15.13 Gesture Study: Questionnaires

15.13.1 Pre-study questionnaire

Please fill out the following questions as completely as possible.

Age (in years): _____
Gender: M F
Department (e.g. Chemistry): _____
Position:

- ☐ Undergraduate student
- ☐ Graduate student
- ☐ Research Associate
- ☐ Postdoctoral Fellow
- ☐ Faculty
- ☐ Staff
- ☐ Other_____

How long in position (e.g. 10 years): _____

Area of Research (e.g. Bioinformatics):

_____

How often do you attend research presentations (e.g. 1/month, 3/week, never)?

_____

How often do you give research presentations (e.g. 1/month, 3/week, never)?

_____

How do you give presentations (e.g. speaking, powerpoint, whiteboard, overhead)?

_____

How often do you use computers (e.g. 2 hours per day, once a week)?

_____

15.13.2 Mid-study questionnaire

Please answer the following questions as completely as possible.

What are the five types of gas that exist in our atmosphere as described in the video?

_____
_____
_____

Which of these gases absorb long-wavelength radiation?

_____
_____
_____

According to the presenter, what will happen to the Earth's temperature if there is too little long-wavelength absorbing gas in the atmosphere?

_____
_____
_____

According to the presenter, what will happen to the Earth's temperature if there is too much long-wavelength absorbing gas in the atmosphere?

_____
_____
_____

15.13.3 Post Study Questionnaire

Please answer the following questions as completely as possible.

What is the goal of the four-quadrant diagram that the presenter uses in his presentation?

What do the rows in the diagram represent?

What do the columns in the diagram represent?

According to the presenter, what are the potential risks of taking action against global warming? Which of these risks did the presenter list in the four quadrant diagram?

According to the presenter, what are the potential risks of not taking action against global warming? Which of these risks did the presenter list in the four quadrant diagram?

According to the presenter, which of the rows in the diagram is most likely to occur? What rationale does the presenter use to justify this position?

According to the presenter, which of the columns in the diagram has the most significant risk? What rationale does the presenter use to justify this position?

According to the presenter, what is the key unknown in trying to understand what action we should take to deal with global warming?

15.14 Gesture Study: Detailed Experimental Analysis

Chapter 9 looks at measures that are consistent across all subjects, as it only considers those scenes in which there is no experimental intervention. In this appendix, we provide a detailed analysis of the impact of our experimental interventions on gaze fixation. In particular, we consider the visibility of facial features and gesture as the independent variables. Note that this appendix contains only our highly detailed statistical analyses. The majority of these results are presented in summary form in Chapter 10 along with our high-level analysis of the outcomes of this study. Those readers that are primarily interested in the high-level results from this study rather than the detailed statistical analyses are referred to Chapter 10.

We perform a detailed analysis of two types of measures in this chapter. First, we consider measures of the collaboration process (Section 15.14.1). These are the measures that help us to determine whether researchers attend to artifacts and whether gestures assist in drawing attention to those artifacts. Ultimately, a scientific presentation is about communicating concepts and information clearly. The analysis of task measures (Section 15.14.2) helps to determine whether gesture and facial expression visibility have an impact on the understanding of the research presentation.

15.14.1 Measures of process

Our measures of process consider the impacts of gesture and facial expression visibility on the fixation times that fall within our AOIs. In particular, we measure the eye fixation times that occur in our scenes. We are concerned with four main communication events: emphatic, implicit, explicit, and manipulation communication events. The goal of our study is to measure the impact of our experimental intervention on eye fixations in the relevant EmphaticGesture, ImplicitPointArtifact, ExplicitPointArtifact, and ArtifactManip AOIs. For more information on the definition of the communication events see Section 7.1.3 of our Ethnography. For more details on the definitions of the AOI types see Section 8.3.1.3.

In the following sections, we divide up the presentation of our statistical analysis based on the scenes defined by the communication events (emphatic, implicit, explicit, and manipulation) that occur in those scenes. We then consider the impact of gesture and facial feature visibility on three measures, the amount of fixation time spent in artifact

AOIs, the amount of fixation time spent in facial expression AOIs, and the total amount of time spent fixated in any type of AOI in the scenes under consideration.

15.14.1.1 Emphatic gesture events

Emphatic gestures are those gesture/utterance pairs that are used for emphasis but do not play a role in artifact interaction. An example of such a gesture, including a hot-spot analysis for the YGNH condition is shown in Figure 72. There are nine scenes in Act 4 in which emphatic gestures occur. We measure fixation time within the EmphaticGesture AOI for each subject in each of the nine scenes that contain an emphatic gesture. We aggregate the EmphaticGesture fixation times for each subject across emphatic gesture scenes. We then analyze the total fixation times per subject based on our experimental conditions (gesture and facial expression visibility).

Dependent Variable:EmphaticGesture

Gesture	Head	Mean	Std. Deviation	N
N	N	243.56	344.976	9
	Y	227.13	230.886	8
	Total	235.82	287.920	17
Y	N	1481.80	1031.643	10
	Y	247.40	326.814	10
	Total	864.60	977.607	20
Total	N	895.26	994.245	19
	Y	238.39	280.362	18
	Total	575.70	801.353	37

Table 20: Statistics for total fixation time (ms) within EmphaticGesture AOIs

The descriptive statistics for the total fixation time (measured in ms) within EmphaticGesture AOIs are provided in Table 20. The table shows a clear difference between the mean and standard deviation for the YGNH condition (see Figure 73). Similar differences can be seen when considering the percentage of overall scene time spent fixating on EmphaticGesture AOIs, with percentages of 1%, 1%, 7%, and 1% of the total scene time for the YGYH, NGNH, YGNH, and NGYH conditions respectively.

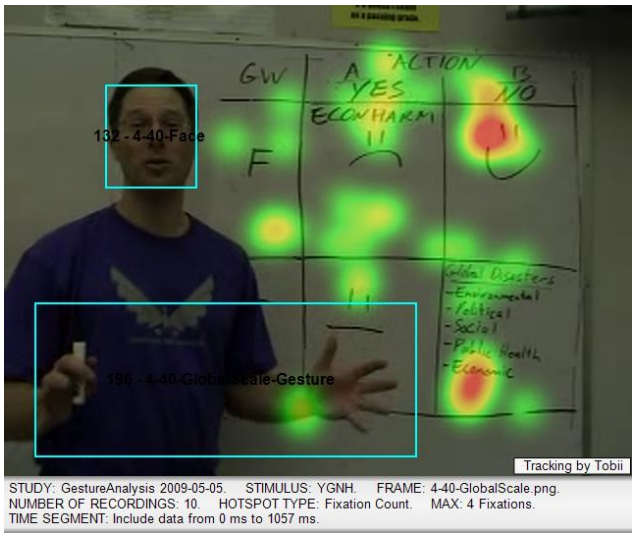


Figure 72: Hot-Spot analysis for EmphaticGesture in the YGNH condition

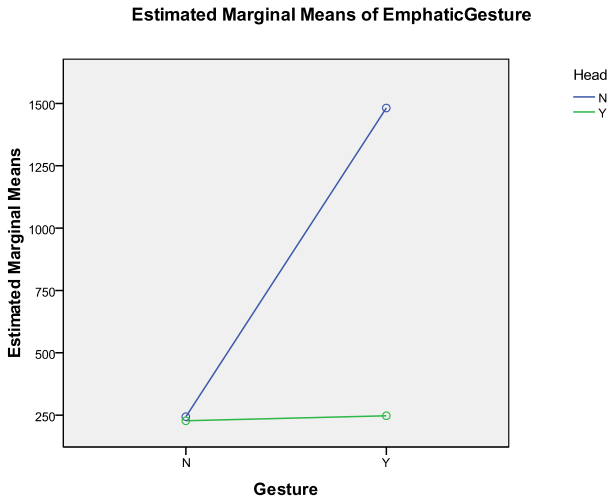


Figure 73: Means for EmphaticGesture

The total EmphaticGesture fixation time in these scenes is normally distributed across all conditions (Kolmogorov-Smirnov Z test giving p = 0.627, 0.673, 0.944, and 0.863 for the YGYH, NGNH, YGNH, and NGYH conditions respectively) but the variances are not homogeneous (Levene's F = 10.459, p = 5.478×10^{-5}). We therefore use a one-way ANOVA to test the equality of the means across the conditions. Our ANOVA shows that there is a significant difference between the means across the four conditions (F = 10.432, p = 5.59×10^{-5}).

We perform a Tamhane post-hoc analysis of the four conditions (error variances are unequal). The mean EmphaticGesture AOI fixation time of the YGNH condition is

significantly higher than the means of the YGYH, NGNH, and NGYH conditions, ($p = 0.025$, $p = 0.025$, and $p = 0.023$ respectively). None of the other conditions have means that are significantly different.

Based on this analysis, there are two clear results. First, that gesture visibility has a significant impact on the level of attention paid to EmphaticGesture AOIs, but this difference only occurs when facial features are not visible (YGNH condition). Second, that although there is a significant difference in the YGNH condition, emphatic gesture is not attended to at the same level as that of the other AOI types that we have considered thus far. We consider these results in the context of the remainder of our analysis in Chapter 10.

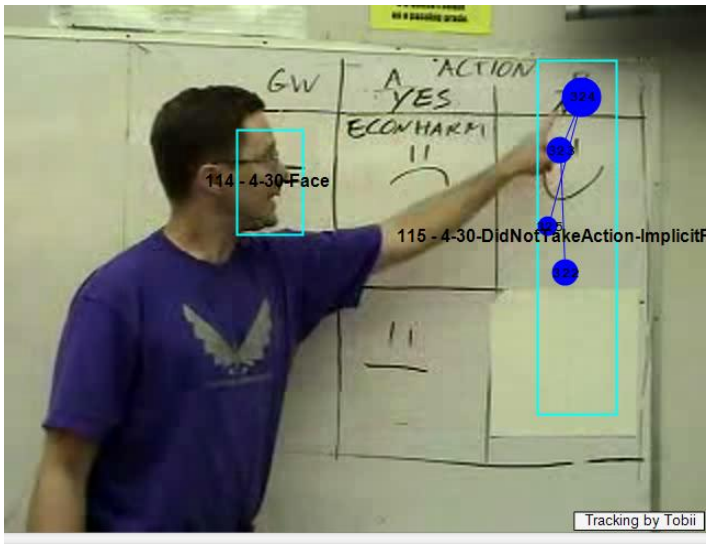


Figure 74: An implicit artifact event scene with relevant AOIs

15.14.1.2 Implicit artifact events

We now consider the scenes in which implicit artifact events occur. Recall that implicit artifact events are those in which an artifact pointing gesture occurs at the same time as an utterance that refers to an artifact on the screen without deixis. In such a gesture/utterance pair, the referent artifact is implicit in the utterance (participants can infer the artifact without the artifact pointing gesture). For example, the speaker stating "… we didn't take action…", while referring to the column in the Pascal's Wager diagram that represents taking no action, would be an implicit artifact event, as study participants can infer from the utterance and the diagram which column is being

discussed (see Figure 74). There are 36 scenes in which implicit artifact events occur and are covered by our experimental intervention (Acts 2, 4, and 6).

We measure fixation time within the ImplicitPointArtifact AOIs[9] for each subject in each scene that contains an implicit artifact event. We aggregate the fixation times for each subject across these scenes. We then analyze the total fixation times across the subjects in our experimental conditions (gesture and facial expression visibility). We consider similar measures for fixation time in ImplicitPointArtifactPost AOIs, FacialFeature AOIs, and all AOI types in these scenes as well.

15.14.1.2.1 ImplicitPointArtifact AOI Analysis

The Kolmogorov-Smirnov Z test of normality for the YGYH, NGNH, YGNH, and NGYH conditions on the ImplicitPointArtifact AOI fixation times indicate that all measures are approximately normal ($p = 0.979$, 0.571, 0.989, and 0.944). Levene's test indicates that the variances across the conditions are homogeneous ($F = 2.863$, $p = 0.052$). The means and standard deviations of the ImplicitPointArtifact AOI fixation times are shown in Table 21.

Dependent Variable:ImplicitPointArtifactAct246

Gesture	Head	Mean	Std. Deviation	N
N	N	10882.00	3024.292	9
	Y	7301.25	1516.043	8
	Total	9196.94	2995.449	17
Y	N	16819.10	4857.768	10
	Y	12788.70	3419.947	10
	Total	14803.90	4581.807	20
Total	N	14006.79	5013.989	19
	Y	10349.83	3874.385	18
	Total	12227.73	4805.434	37

Table 21: Descriptive statistics for ImplicitPointArtifact fixation time (ms)

Since we have data that is approximately normally distributed and the error variances are approximately equal, we use a two-way ANOVA (gesture and facial expression) with two levels for each factor (visible and not visible) to test for differences in the mean fixation time across the scenes. There is a statistically significant facial feature visibility

[9] Recall that the ImplicitPointArtifact AOI captures the region of the screen in which the artifact resides, not where the gesture took place. That is, it is an **artifact** AOI, not a gesture AOI.

effect (F = 10.773, p = 0.002, p < 0.05). There is also a statistically significant gesture visibility effect (F = 24.272, p = 2.29 x 10^{-5}, p < 0.05). There is no interaction effect between gesture and facial feature visibility (F = 0.038, p = 0.847). The estimated means showing the effects can be seen in Figure 75.

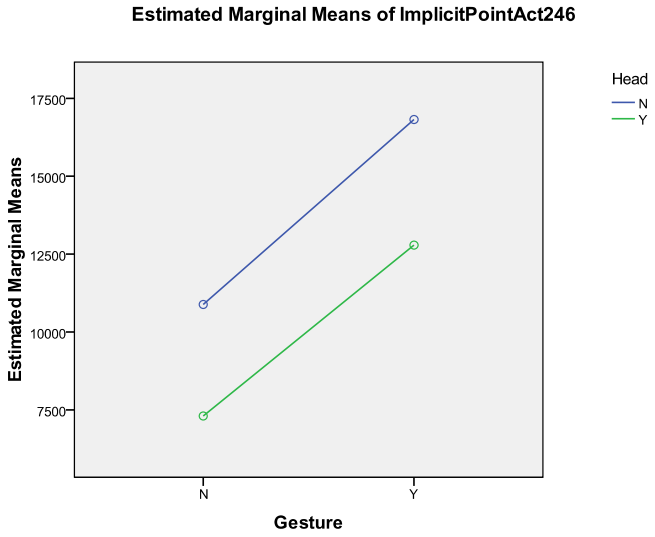


Figure 75: Means of ImplicitPointArtifact fixation times (ms) in implicit artifact scenes

15.14.1.2.2 ImplicitPointArtifactPost AOI analysis

We now consider the ImplicitPointArtifactPost AOIs across the relevant scenes. Recall that such AOIs are identical to the ImplicitPointArtifact AOIs that were referred to by artifact gestures but occur in the scenes immediately following the implicit artifact event scenes. ImplicitPointArtifactPost AOIs are used to capture gaze fixations on artifacts that are associated with the gesture but occur after the gesture has finished (and therefore a new scene has begun). For example, gaze lingering on an artifact after the pointing gesture that brought our attention to that artifact would be captured by such an AOI.

These AOIs also allow us to capture gaze fixations on an artifact in experimental no gesture (NG) conditions. This is of particular relevance in implicit artifact events, where the utterance implies the referent artifact at the same time as the gesture is made (in the example above, the “do not take action” column of the diagram). In the conditions where gesture is not visible, subjects have to utilize the utterance and a visual search to find the artifact. Without a visible artifact gesture, this search may take longer than the scene duration (which is defined by the gestural action). The cognitive psychology literature

indicates that focusing of attention after an artifact has been identified by an utterance takes on average 250 ms [TSE+95]. As a result, valid fixations on the ImplicitPointArtifact AOI may occur outside of the scene. It is important that we include these fixations, as they are a result of the ImplicitPointArtifact action that defines the scene and our analysis. Using the same ANOVA procedure as above, we determine that there is a statistically significant facial feature visibility effect ($F = 15.163$, $p = 4.54 \times 10^{-4}$) but there is no statistically significant gesture visibility effect ($F = 0.779$, $p = 0.384$) or interaction effect ($F = 0.199$, $p = 0.658$). Note that facial feature visibility decreases the fixation time in the ImplicitPointArtifactPost AOIs. The estimated means are shown in Figure 76.

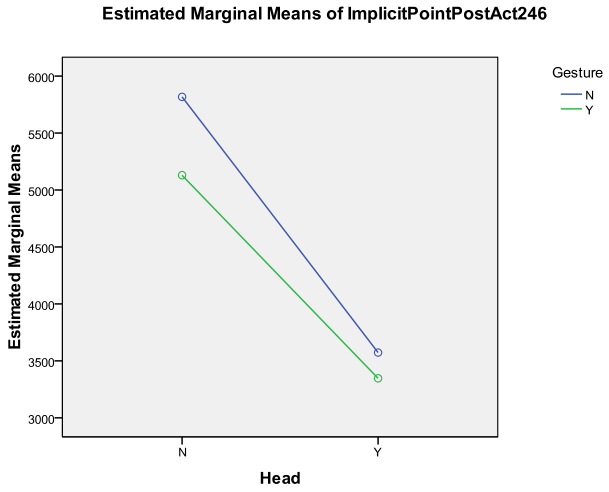


Figure 76: Means of ImplicitPointArtifactPost fixation times (ms) in implicit artifact scenes

15.14.1.2.3 FacialFeature AOI analysis

We also consider the fixation time in the Facial Expression AOIs across the implicit artifact event scenes. The FacialFeature AOI fixation times indicate are approximately normal (Kolmogorov-Smirnov Z, $p = 0.970$, 0.262, 0.746, and 0.970). The variances across the conditions are not homogeneous (Levene's $F = 13.750$, $p = 5.52 \times 10^{-6}$). As one might expect (see Figure 77), there is a dramatic difference between the visible (YH) and non-visible (NH) facial feature conditions (given that one can not see the face in the non-visible conditions). A one-way ANOVA indicates that there is a significant difference in the means across the conditions ($F = 78.75$, $p = 4.00 \times 10^{-15}$). A Tamhane

post-hoc test to compare the means (non-homogeneous variances) indicates that the mean FacialFeature AOI fixation time of the NGYH condition is significantly higher than the YGYH, NGNH, and YGNH conditions ($p = 0.018$, $p = 1.86 \times 10^{-5}$, and $p = 2.09 \times 10^{-5}$ respectively). The YGYH mean is also significantly higher than the NGNH ($p = 1.48 \times 10^{-4}$) and YGNH condition ($p = 1.39 \times 10^{-4}$). The two NH conditions are not significantly different from each other ($p = 0.772$).

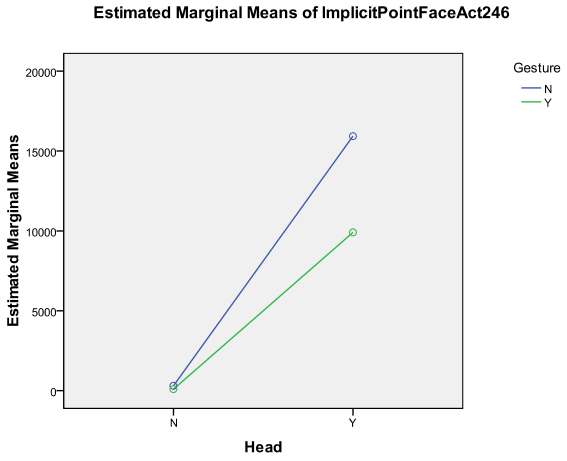


Figure 77: Means of FacialFeature AOI fixation times (ms) in implicit artifact scenes

15.14.1.2.4 Total AOI fixation time analysis

Lastly, we consider the total fixation time across all AOIs for implicit artifact event scenes. This allows us to get an overall understanding of how our experimental interventions affect the overall fixation time spent in AOIs. Overall, total fixations time spent within AOIs across the four conditions account for 47%, 45%, 36%, and 24% of the total scene time for the YGYH, NGYH, YGNH, and NGNH conditions respectively. As in Section 8.7.1.2, we see that our AOIs capture a significant percentage of the total scene time, even in the least visually rich NGNH condition.

The total AOI fixation time across our conditions is approximately normally distributed (Kolmogorov-Smirnov Z, $p = 0.994$, 0.684, 0.766, 0.613) with approximately homogeneous variances (Levene's $F = 1.32$, $p = 0.284$). We therefore use a two-way ANOVA to test for differences in the mean total fixation time across all AOIs and across all implicit artifact event scenes. There is a statistically significant facial feature visibility effect ($F = 29.42$, $p = 5.28 \times 10^{-6}$). There is also a statistically significant gesture effect (F

= 5.474, p = 0.026). Although there is a moderate interaction effect between gesture and facial feature visibility, this effect is not statistically significant at the $\alpha = 0.05$ level (F = 3.424, p = 0.073). The estimated means showing the overall AOI fixation effects can be seen in Figure 78. We perform a post-hoc Tukey HSD multiple comparison analysis on the four conditions. The average overall fixation time of the NGNH condition is smaller than the YGYH, YGNH, and NGYH conditions ($p = 1.92 \times 10^{-5}$, p = 0.024, and $p = 1.23 \times 10^{-4}$). The YGYH, YGNH, and NGYH are not statistically different at a $\alpha = 0.05$ level, although there is a moderately significant difference in YGYH and YGNH conditions (p = 0.058).

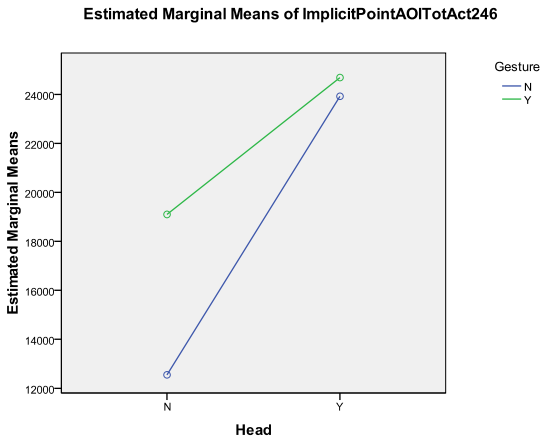


Figure 78: Means for total AOI fixation time (ms) in implicit artifact scenes

15.14.1.3 Explicit artifact events

Explicit artifact events are one of the most important types of gestures to consider in the context of distributed, artifact-centric, scientific collaboration. An explicit artifact event consists of an utterance and a gesture occurring at the same time, with no information in the utterance about the referent artifact (deixis). We call these actions explicit gestures because without the explicit pointing gesture to the referent artifact the gesture/utterance pair is meaningless. That is, the referent is completely dependent on the context of the discussion, and without the explicit artifact gesture the observer has no idea to which artifact the speaker is referring. In Chapter 7 we show that explicit artifact events are utilized extensively in both collocated and distributed scientific collaboration. We also show that much of the gestural information is lost when those collaborations take

place at a distance. This is of particular importance for explicit (deictic) gestures, as they lose all communicative meaning if the artifact gesture is not visible. This section of the study provides quantitative information about how artifacts that are referred to by explicit artifact events are attended to during scientific communication.

In this section, we consider the scenes in which explicit artifact events occur. There are 12 scenes spread across Act 4 and Act 6 that contain explicit artifact events. We measure fixation time within the ExplicitPointArtifact and FacialFeature AOIs in each scene. We also measure the total fixation time in all AOIs. As in previous sections, we aggregate fixation times across all of the explicit artifact event scenes for each subject and then analyze the aggregate fixation times across the conditions.

	Explicit YGYH	Explicit NGNH	Explicit YGNH	Explicit NGYH	Face YGYH	Face NGNH	Face YGNH	Face NGYH	Total YGYH	Total NGNH	Total YGNH	Total NGYH
Kolmogorov-Smirnov Z	.496	.961	.509	.468	1.074	.956	1.290	.461	.536	.543	.459	.441
Asymp. Sig. (2-tailed)	.966	.315	.958	.981	.199	.321	.072	.984	.936	.930	.984	.990

Table 22: Kolmogorov-Smirnov Z test for normality in explicit artifact scenes

Dependent Variable:ExplicitPointAct46

Gesture	Head	Mean	Std. Deviation	N
N	N	607.00	678.515	9
	Y	713.37	478.951	8
	Total	657.06	577.535	17
Y	N	2168.60	1118.619	10
	Y	2492.80	847.129	10
	Total	2330.70	979.955	20
Total	N	1428.89	1213.262	19
	Y	1701.94	1141.132	18
	Total	1561.73	1170.497	37

Table 23: ExplicitPointArtifact AOI fixation times (ms) in explicit artifact scenes

15.14.1.3.1 ExplicitPointArtifact AOI analysis

We now consider the fixation times in the ExplicitPointArtifact AOIs across the explicit point scenes. The Kolmogorov-Smirnov Z test for normality on the ExplicitPointArtifact, FacialFeature, and total AOI fixation times shows that all distribution are approximately normal (Table 22). The descriptive statistics for ExplicitPointArtifact AOIs are given in Table 23. Levene's test of equality of variances ($F = 2.599$, $p = 0.069$) shows that the error variances are approximately equal. A two-way

ANOVA indicates that there is a statistically significant gesture visibility effect ($F = 36.709$, $p = 8.12 \times 10^{-7}$) but there is no head visibility ($F = 0.610$, $p = 0.440$) or interaction ($F = 0.156$, $p = 0.695$) effect in the mean fixation time in ExplicitPointArtifact AOIs. Gesture visibility has a clear impact on artifact attention in scenes where explicit artifact events occur.

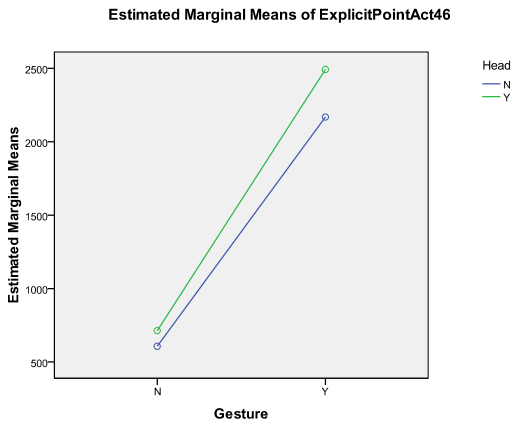


Figure 79: Means of ExplicitPointArtifact fixation times (ms) in explicit artifact scenes

Dependent Variable:ExplicitPointFaceAct46

Gesture	Head	Mean	Std. Deviation	N
N	N	110.89	181.568	9
	Y	3128.13	1321.835	8
	Total	1530.76	1786.256	17
Y	N	65.80	120.216	10
	Y	592.50	762.634	10
	Total	329.15	596.112	20
Total	N	87.16	149.709	19
	Y	1719.44	1645.677	18
	Total	881.24	1405.061	37

Table 24: FacialFeature AOI fixation time (ms) in explicit artifact scenes

15.14.1.3.2 FacialFeature AOI analysis

We also consider the fixation time in the FacialFeature AOIs across the scenes with explicit artifact events. The descriptive statistics for fixation time within FacialFeature AOIs are shown in Table 24. As expected, in the non-visible facial feature (NH) conditions, almost no time is spent fixated on the FacialFeature AOI. In the visible face conditions, the average FacialFeature AOI fixation time is higher when there is no

gesture visible (see Figure 80). Since the variances are not homogeneous (Levene's F = 12.643, $p = 1.15 \times 10^{-5}$), we utilize a one-way ANOVA with the Tamhane post-hoc test to test for equality of the mean fixation time across the four conditions.

Our analysis reveals that there is a statistically significant difference in the means (F = 32.775, $p = 5.19 \times 10^{-10}$). The Tamhane post-hoc pair-wise test shows that the NGYH condition is significantly different from the YGYH (p = 0.004), NGNH (p = 0.002), and YGNH (p = 0.002) conditions. None of the other conditions differ significantly (α = 0.05). It is worth noting that the mean fixation time on FacialFeature AOIs in the YGYH condition is not significantly different than the non-visible facial feature conditions in which there is no facial expression visible at all. This implies that subjects do not attend to facial features often when explicit artifact events occur.

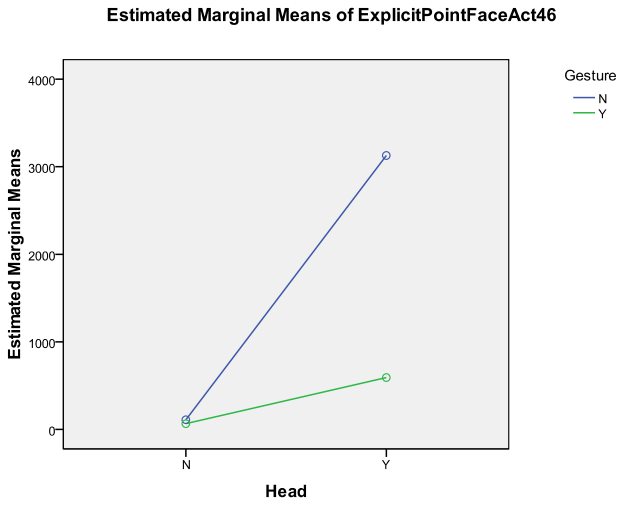


Figure 80: Means of FacialFeature fixation times (ms) in explicit artifact scenes

Dependent Variable:ExplicitPointAOITotalAct46

Gesture	Head	Mean	Std. Deviation	N
N	N	1475.78	868.656	9
	Y	4514.25	1454.261	8
	Total	2905.65	1935.559	17
Y	N	3183.80	1359.652	10
	Y	4455.50	1358.160	10
	Total	3819.65	1474.795	20
Total	N	2374.74	1423.871	19
	Y	4481.61	1359.517	18
	Total	3399.70	1739.606	37

Table 25: Statistics for total fixation time (ms) in explicit artifact scenes

15.14.1.3.3 Total AOI fixation time analysis

In order to understand the amount of overall fixation time spent in AOIs, we consider the total AOI fixation time for each subject across the explicit artifact event scenes. The percentage of fixation time spent in AOIs over all subjects and all explicit artifact scenes are 28%, 26%, 18%, and 5% for the YGYH, NGYH, YGNH, and NGNH conditions respectively. The facial features visible (YH) conditions have the highest mean fixation times. The YGNH condition also has a high mean fixation time. The NGNH condition, not surprisingly, has the lowest mean overall fixation time. Levene's test shows that the variances across condition are approximately homogeneous ($F = 0.689$, $p = 0.565$). A two-way ANOVA results in a statistically significant facial feature visibility effect ($F = 25.99$, $p = 1.38 \times 10^{-5}$). There is also a statistically significant interaction effect ($F = 4.368$, $p = 0.044$). There is a moderately significant gesture visibility effect ($F = 3.806$, $p = 0.06$). The estimated means showing the overall AOI fixation effects can be seen in Figure 81. We perform a post-hoc Tukey HSD comparison analysis on the four conditions. The average overall fixation time of the NGNH condition is smaller than the YGYH ($p = 8.63 \times 10^{-5}$), YGNH ($p = 0.032$), and NGYH ($p = 1.46 \times 10^{-4}$) conditions. The YGYH, YGNH, and NGYH means are not statistically different at a α level of 0.05.

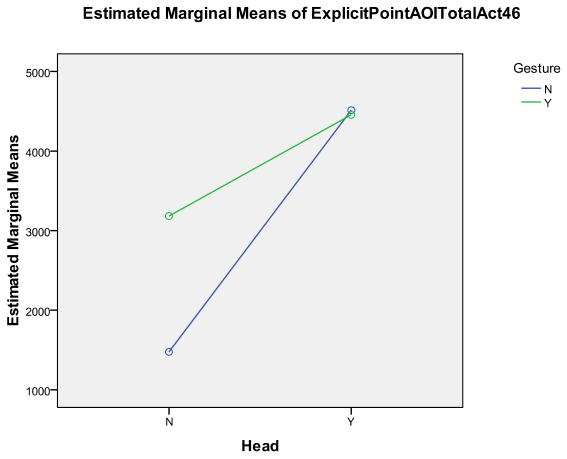


Figure 81: Fixation time in all AOIs (ms) in explicit artifact scenes

15.14.1.4 Artifact manipulation

An important aspect of scientific communication is the direct manipulation of the artifacts that are part of the collaboration. The results from our ethnography (Section 7.3) show that artifact manipulation is a common operation across a wide range of scientific communication scenarios. This includes selecting artifacts, annotating artifacts (circling or underlying artifacts), transforming artifacts (moving or deleting artifact components), or extending artifacts (adding new artifacts). In the presentation used in this study, artifact manipulation is primarily carried out by the presenter by either writing or drawing on the diagram. Manipulation operations that are performed are selecting/annotating (circling a part of the diagram), transforming (erasing parts of the diagram), and extending (adding to the diagram through writing). We consider these actions as a type of gestural manipulation, as the presenter's hand is used to perform the writing action.

Our experimental intervention of gesture visibility allows us to consider the effects of gesture on attention to artifact manipulation. We also are able to determine the effect of facial feature visibility on the level of attention to artifact manipulation. There are 11 artifact manipulation scenes in the presentation, nine in Act 4 and two in Act 6. We measure fixation time within the ArtifactManip and FacialFeature AOIs. We also measure the total fixation time in all AOIs. As in previous sections, we then aggregate fixation times across all of the artifact manipulation scenes and analyze the effect of our interventions across conditions.

The artifact manipulation scenes result in one of the highest percentages of fixation time within AOIs (percentage of overall scene time), with 57% of the overall scene time spent as a fixations within one of the scene AOIs. In the YGYH condition over 72% of the total scene time is spent as a fixation within the scene AOIs. Of particular importance, the ArtifactManip AOIs account for a very larger proportion of the total scene time, with 52% of the overall time spent fixated on the ArtifactManip AOIs and 70%, 37%, 48%, and 49% of the total scene time spent within ArtifactManip AOIs in the YGYH, NGNH, YGNH, and NGYH conditions respectively. Clearly, artifact manipulation in this form is a focal point for subjects in this study. We explore this in more detail below.

15.14.1.4.1 ArtifactManip AOI analysis

	ArtifactManip				Face				AOI Total				ArtifactManipPost			
	YGYH	NGNH	YGNH	NGYH	YGYH	NGNH	YGNH	NGYH	YGYH	NGNH	YGNH	NGYH	YGYH	NGNH	YGNH	NGYH
Z	.630	.463	.806	.572	.732	.389	.487	.528	.613	.580	.711	.818	.498	.552	.608	.570
P	.822	.983	.535	.899	.658	.998	.972	.943	.846	.890	.693	.516	.965	.921	.854	.901

Table 26: Kolmogorov-Smirnov Z test in artifact manipulation scenes

Dependent Variable:ArtifactManipAct46

Gesture	Head	Mean	Std. Deviation	N
N	N	7071.89	2525.191	9
	Y	9287.00	2274.338	8
	Total	8114.29	2598.105	17
Y	N	9103.40	2433.422	10
	Y	13290.20	3051.232	10
	Total	11196.80	3439.174	20
Total	N	8141.11	2623.142	19
	Y	11511.00	3353.868	18
	Total	9780.51	3415.842	37

Table 27: Descriptive statistics for ArtifactManip fixation times (ms)

Our tests for normality show that our measures across all AOI types are approximately normal (Table 26). Levene's test on the ArtifactManip AOI fixation times shows that the variances are approximately homogeneous ($F = 0.451$, $p = 0.718$). As described above, the fixation times spent within the ArtifactManip AOIs account for a large proportion of the overall scene time. The descriptive statistics for this measure are given in Table 27. From this table, we see that the YGYH condition has the largest mean and the NGNH condition has the lowest mean, with the NGYH and YGNH means being relatively close.

A two-way ANOVA indicates that there are statistically significant gesture ($F = 12.285$, $p = 0.001$) and facial feature ($F = 13.826$, $p = 0.001$) visibility effects. There is no interaction ($F = 1.311$, $p = 0.260$) effect in the mean fixation time in ArtifactManip AOIs. The estimated means for the ArtifactManip fixation time AOIs can be seen in Figure 82. A Tukey HSD post-hoc analysis shows that the YGYH condition is significantly higher than the NGNH ($p = 5.99 \times 10^{-5}$), YGNH ($p = 0.006$), and NGYH ($p = 0.014$) conditions. The NGNH, YGNH, and NGYH conditions are not significantly different at an α level of 0.05.

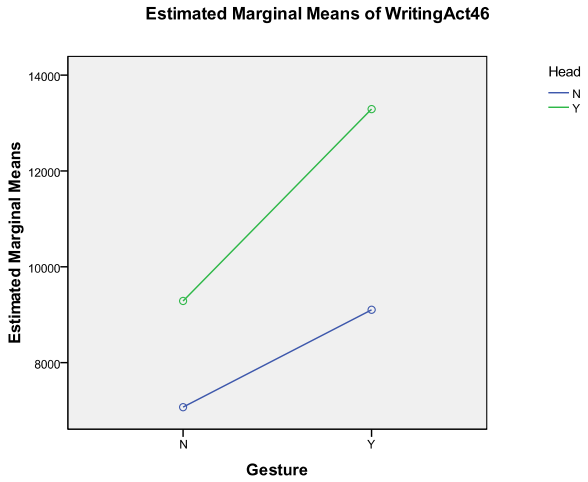


Figure 82: Means of ArtifactManip fixation times (ms) in artifact manipulation scenes

Dependent Variable:ArtifactManipPostAct46

Gesture	Head	Mean	Std. Deviation	N
N	N	5337.33	1972.755	9
	Y	3172.88	1149.676	8
	Total	4318.76	1940.171	17
Y	N	4675.30	1927.629	10
	Y	3612.30	1489.617	10
	Total	4143.80	1763.105	20
Total	N	4988.89	1924.289	19
	Y	3417.00	1330.217	18
	Total	4224.19	1822.481	37

Table 28: Statistics for ArtifactManipPost fixation time (ms) in artifact manipulation scenes

15.14.1.4.2 ArtifactManipPost AOI analysis

Artifact manipulation scenes draw a significant amount of fixation time within AOIs, with up to 72% of the total scene time spent within AOIs in these scenes. The ArtifactManip AOIs in particular are the focus of much of this attention. In pre-study testing, it was noted that fixations often continued after the manipulation process, and therefore the artifact manipulation scene, ended. We therefore created ArtifactManipPost AOIs in the scenes immediately following the ArtifactManip scenes. For example, gaze lingering on a written artifact after the artifact manipulation finished would be captured by ArtifactManipPost AOIs.

Levene's test shows that fixations in ArtifactManip AOIs have homogeneous variances ($F = 0.850$, $p = 0.476$). A two-way ANOVA analysis of the ArtifactManipPost fixations show that there is a statistically significant facial feature visibility effect ($F = 8.403$, $p = 0.007$) but there is no statistically significant gesture visibility effect ($F = 0.040$, $p = 0.843$) or interaction effect ($F = 0.979$, $p = 0.330$). The estimated means are shown in Figure 83. A Tukey HSD post-hoc analysis shows that the NGNH ArtifactManipPost AOI fixations are moderately higher than those of the NGYH condition, with no significant differences between other pairs. This implies that although subjects attend to ArtifactManipPost AOIs, when facial features are visible the level of attention is reduced.

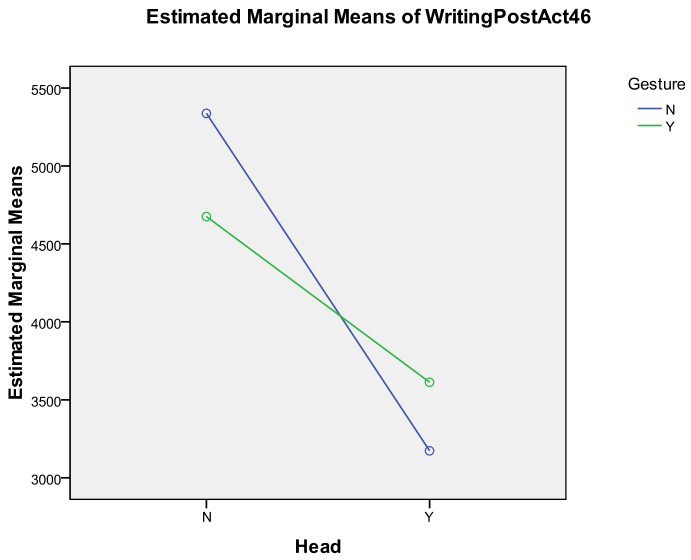


Figure 83: Means of ArtifactManipPost fixation times (ms) in artifact manipulation scenes

Dependent Variable:ArtifactManipFaceAct46

Gesture	Head	Mean	Std. Deviation	N
N	N	423.11	343.064	9
	Y	1418.63	761.697	8
	Total	891.59	758.296	17
Y	N	308.90	260.868	10
	Y	179.50	187.037	10
	Total	244.20	230.678	20
Total	N	363.00	299.611	19
	Y	730.22	811.689	18
	Total	541.65	625.001	37

Table 29: Statistics for FacialFeature AOI fixation time (ms) in artifact manipulation scenes

15.14.1.4.3 FacialFeature AOI analysis

We also consider the FacialFeature AOIs across the artifact manipulation scenes. Given that in some conditions there is no visible facial feature, it is to be expected that there are dramatic differences in the means across the conditions. Levene's test shows that the variances are not homogeneous across our conditions ($F = 5.029$, $p = 0.006$). The ANOVA indicates the means across our conditions are not equal ($F = 15.085$, $p = 2.36 \times 10^{-6}$). A Tamhane post-hoc analysis shows that the NGYH mean is significantly higher than the YGYH ($p = 0.013$), NGNH ($p = 0.043$), and YGNH ($p = 0.024$) means (Figure 84). This implies that when gesture is visible, FacialFeature AOIs are not attended to significantly more than when there is no facial features visibility at all.

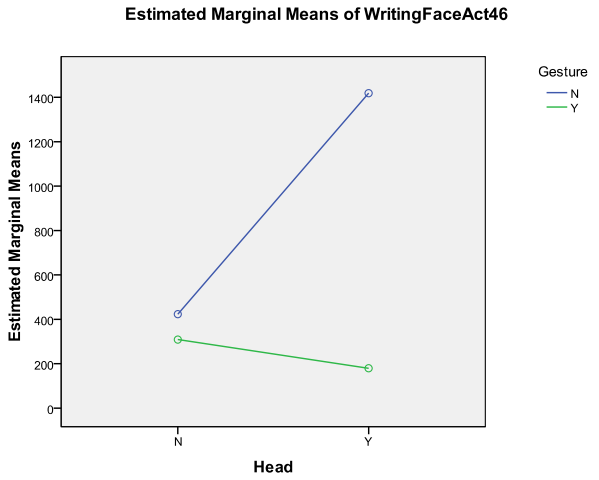


Figure 84: Means of FacialFeature fixation times (ms) in artifact manipulation scenes

Dependent Variable:ArtifactManipAOITotalAct46

Gesture	Head	Mean	Std. Deviation	N
N	N	8408.22	3011.949	9
	Y	11239.13	2281.431	8
	Total	9740.41	2989.052	17
Y	N	9861.20	2784.550	10
	Y	13725.20	3171.851	10
	Total	11793.20	3516.740	20
Total	N	9172.95	2909.356	19
	Y	12620.28	3014.177	18
	Total	10850.03	3402.022	37

Table 30: Statistics for Total AOI fixation time (ms) in artifact manipulation scenes

15.14.1.4.4 Total AOI fixation time analysis

As with the other scene types, we complete our analysis of the artifact manipulation scenes by considering the total fixation time across all AOIs in these scenes. Levene's test of equality of variance indicates that the variances are approximately equal across the four conditions ($F = 0.388$, $p = 0.763$). A two-way ANOVA indicates that there is a significant gesture visibility effect ($F = 4.359$, $p = 0.045$) and a significant facial feature visibility effect ($F = 12.592$, $p = 0.001$). There is no interaction effect between gesture and facial feature visibility ($F = 0.3$, $p = 0.588$). The YGYH condition has the largest mean fixation time, with the NGNH condition having the lowest. Performing a Tukey HSD mean differences test on the conditions individually shows that the YGYH condition is significantly different than the NGNH and YGNH ($p = 0.002$ and $p = 0.024$ respectively) but not the NGYH condition ($p = 0.276$). There are no other statistically significant pair-wise differences.

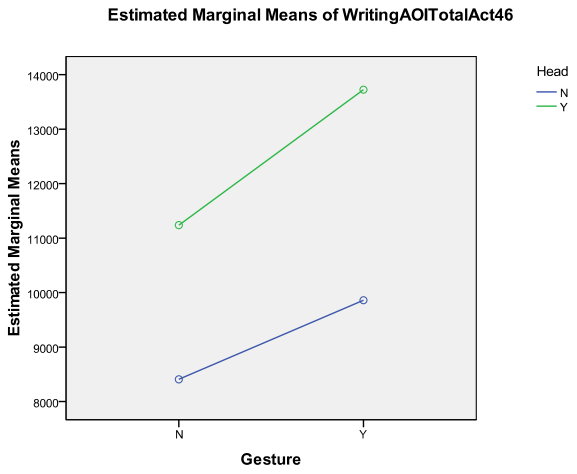


Figure 85: Estimated means of total AOI fixation times (ms)

15.14.2 Task Measures

Section 10.1 and Section 15.14.1 both consider measures of the communication process used in scientific collaboration. That is, we measure and analyze how the subjects attend to the human communication channels presented during a scientific presentation. We also need to understand the impact of our experimental intervention on the task. The post-study questionnaire is designed to determine whether our experimental interventions affect the understanding of the topic of the presentation. Recall that the questionnaire was designed to test three main aspects of our study participant's understanding:

- Understanding about the artifact: We test our participant's understanding of the structural nature of the Pascal's Wager diagram and in particular the roles the rows and columns play in the presentation (Question 2 and 3).
- Understanding about information: We test our participant's understanding of the information presented using the Pascal's Wager diagram and in particular the recollection of specific facts that were presented during the presentation (Question 4 and 5).
- Understanding about the argument: We test our participant's understanding of the argument being made by the presenter and in particular the participant's recollection of several key facts that the presenter used to make his argument (Question 6 and Question 7).

Question 1, Question 8, and Question 9 were open ended questions that provide us with qualitative information about participants understanding of the presentation. These questions do not have correct answers, and were therefore not scored for correctness. As such, we do not discuss them in our quantitative analysis below.

It is also worth pointing out that the number of participants in our questionnaire analysis is larger than in our analysis of attention to artifact AOIs. Recall that we had problems with the Tobii eye tracking system, and it was therefore necessary to discard data from those participants. Since we do not need eye tracking data for questionnaire scores, we use all study participants in the analysis below.

15.14.2.1 Overall scores: Q2 – Q7

OverallScore

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval		Minimum	Maximum
					Lower Bound	Upper Bound		
NGNH	14	10.14	1.657	.443	9.19	11.10	7	12
NGYH	14	9.93	2.841	.759	8.29	11.57	1	12
YGNH	11	8.82	2.676	.807	7.02	10.62	4	11
YGYH	11	10.55	1.128	.340	9.79	11.30	9	12
Total	50	9.88	2.228	.315	9.25	10.51	1	12

Table 31: Descriptive statistics for post-study overall score

Initially, we consider the overall scores on Question 2 through Question 7 of all subjects across our four conditions. Note that this is only considering the first part of the answers on those questions that had two part answers (Question 4 through Question 7). Each question is scored out of two marks, for a possible maximum score of 12 points. The descriptive statistics for the total score across the conditions is given in Table 31. The Kolmogorov-Smirnov test for normality reveals that the distribution is non-normal ($Z = 2.208$, $p = 0.001$) and we therefore use the non-parametric Kruskal-Wallis to perform a one-way analysis of variance of the ranks of the samples. The Kruskal-Wallis test gives a Chi-Squared statistic of 3.458 ($p = 0.326$), indicating that there are no significant differences on overall score across the four conditions.

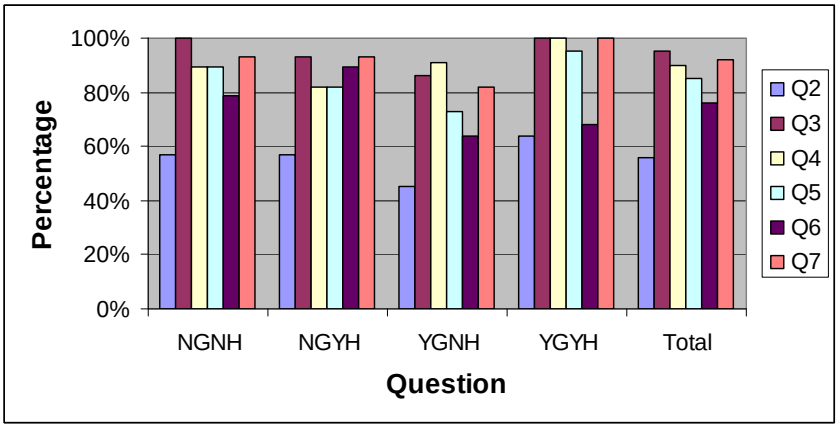


Figure 86: Percentage scores for each question across conditions (Q2 - Q7)

15.14.2.2 Individual scores: Q2 – Q7

Test Statistics – Kruskal Wallis Test

	Q2	Q3	Q4	Q5	Q6	Q7	Overall
Chi-Square	3.693	4.353	2.584	4.051	3.876	4.666	3.458
DF	3	3	3	3	3	3	3
Asymp. Sig.	.297	.226	.460	.256	.275	.198	.326

Table 32: Kruskal-Wallis test statistics for Question 2 through Question 7.

When considering the questions on an individual basis, we see similar results (see Table 32 and Figure 86). In all cases, there is no significant difference in the ranks across the four conditions. Thus at least on the surface it appears we have no significant impact on understanding across our four conditions, despite the sometimes dramatic effect on attention described in Section 15.14.1. Although there is some indication from our pre-study testing that the questionnaire might be slightly simplistic, it was believed that the questions were of sufficient difficulty to reveal the effects of our experimental interventions. Thus, these results are somewhat surprising.

		N	Mean	Std. Dev	Std. Error	95% Con. Int. Lower Bound	95% Con. Int. Upper Bound	Min	Max
Overall	NGNH	14	10.14	1.657	.443	9.19	11.10	7	12
	NGYH	14	9.93	2.841	.759	8.29	11.57	1	12
	YGNH	11	8.82	2.676	.807	7.02	10.62	4	11
	YGYH	11	10.55	1.128	.340	9.79	11.30	9	12
	Total	50	9.88	2.228	.315	9.25	10.51	1	12
Q2	NGNH	14	1.14	.363	.097	.93	1.35	1	2
	NGYH	14	1.14	.363	.097	.93	1.35	1	2
	YGNH	11	.91	.539	.163	.55	1.27	0	2
	YGYH	11	1.27	.467	.141	.96	1.59	1	2
	Total	50	1.12	.435	.062	1.00	1.24	0	2
Q3	NGNH	14	2.00	.000	.000	2.00	2.00	2	2
	NGYH	14	1.86	.535	.143	1.55	2.17	0	2
	YGNH	11	1.73	.647	.195	1.29	2.16	0	2
	YGYH	11	2.00	.000	.000	2.00	2.00	2	2
	Total	50	1.90	.416	.059	1.78	2.02	0	2
Q4	NGNH	14	1.79	.579	.155	1.45	2.12	0	2
	NGYH	14	1.64	.745	.199	1.21	2.07	0	2
	YGNH	11	1.82	.405	.122	1.55	2.09	1	2
	YGYH	11	2.00	.000	.000	2.00	2.00	2	2
	Total	50	1.80	.535	.076	1.65	1.95	0	2
Q5	NGNH	14	1.79	.426	.114	1.54	2.03	1	2
	NGYH	14	1.64	.745	.199	1.21	2.07	0	2
	YGNH	11	1.45	.688	.207	.99	1.92	0	2
	YGYH	11	1.91	.302	.091	1.71	2.11	1	2
	Total	50	1.70	.580	.082	1.54	1.86	0	2
Q6	NGNH	14	1.57	.852	.228	1.08	2.06	0	2
	NGYH	14	1.79	.579	.155	1.45	2.12	0	2
	YGNH	11	1.27	.905	.273	.67	1.88	0	2
	YGYH	11	1.36	.809	.244	.82	1.91	0	2
	Total	50	1.52	.789	.112	1.30	1.74	0	2
Q7	NGNH	14	1.86	.535	.143	1.55	2.17	0	2
	NGYH	14	1.86	.535	.143	1.55	2.17	0	2
	YGNH	11	1.64	.674	.203	1.18	2.09	0	2
	YGYH	11	2.00	.000	.000	2.00	2.00	2	2
	Total	50	1.84	.510	.072	1.70	1.98	0	2

Table 33: Descriptive statistics for Q2-Q7 and Overall score

In exploring these results in more detail (see Table 33 and Figure 86), there are a number of interesting results that are worth discussing. First, for almost all questions the mean score in the YGYH condition is the highest. In addition, for almost all questions the mean score in the YGNH condition is the lowest. Performing a Mann-Whitney U test (because our distributions are non-normal) on these two conditions, we see some mean rank comparisons that are moderately significance ($p < 0.1$) (see Table 34). In particular, the Overall ($p = 0.082$), Question 5 ($p = 0.057$), and Question 7 ($p = 0.069$) scores all result in moderately significant results. In addition, Question 6 shows a similar significance level across the NGYH and YGNH conditions ($p = 0.089$). None of the other pairs result in statistical significance levels below an α level of 0.1. Although these results are not significant at an α level of 0.05, these results do show some trends towards significance across some of the conditions in this study.

	Overall	Q2	Q3	Q4	Q5	Q6	Q7
Mann-Whitney U	35.500	41.500	49.500	49.500	38.000	58.000	44.000
Wilcoxon W	101.500	107.500	115.500	115.500	104.000	124.000	110.000
Z	-1.736	-1.598	-1.447	-1.449	-1.900	-.182	-1.817
Asymp. Sig. (2-tailed)	.082	.110	.148	.147	.057	.856	.069

Table 34: Mann-Whitney U test statistics for the YGYH and YGNH conditions.

15.14.2.3 Question Scoring: Q4a – Q7a

Our early analysis of the questionnaires, combined with our pre-study testing, indicated that our questions may be "too easy". That is, participants were getting high scores in all conditions. This trend can be seen in Figure 86. Our supplementary questions addressed this issue by asking for more detailed information on four of the questions (Q4 through Q7). We denote these questions Q4a through Q7a. These supplementary questions either asked for specific details that were pertinent to the presenter's argument (see Section 8.7.2.3 and Appendix 15.8 for more details on the questionnaire).

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Overall	NGNH	8	15.88	2.357	.833	13.90	17.85	11	18
	NGYH	7	17.86	1.574	.595	16.40	19.31	16	20
	YGNH	5	14.20	4.970	2.223	8.03	20.37	6	19
	YGYH	5	16.80	1.789	.800	14.58	19.02	15	19
	Total	25	16.28	2.937	.587	15.07	17.49	6	20
Q4a	NGNH	8	1.38	.744	.263	.75	2.00	0	2
	NGYH	7	2.00	.000	.000	2.00	2.00	2	2
	YGNH	5	1.80	.447	.200	1.24	2.36	1	2
	YGYH	5	2.00	.000	.000	2.00	2.00	2	2
	Total	25	1.76	.523	.105	1.54	1.98	0	2
Q5a	NGNH	8	.63	.744	.263	.00	1.25	0	2
	NGYH	7	1.43	1.397	.528	.14	2.72	0	4
	YGNH	5	.60	.894	.400	-.51	1.71	0	2
	YGYH	5	1.40	1.140	.510	-.02	2.82	0	3
	Total	25	1.00	1.080	.216	.55	1.45	0	4
Q6a	NGNH	8	1.50	.926	.327	.73	2.27	0	2
	NGYH	7	1.71	.756	.286	1.02	2.41	0	2
	YGNH	5	1.60	.894	.400	.49	2.71	0	2
	YGYH	5	1.60	.894	.400	.49	2.71	0	2
	Total	25	1.60	.816	.163	1.26	1.94	0	2
Q7a	NGNH	8	1.50	.926	.327	.73	2.27	0	2
	NGYH	7	1.71	.756	.286	1.02	2.41	0	2
	YGNH	5	1.60	.548	.245	.92	2.28	1	2
	YGYH	5	1.60	.548	.245	.92	2.28	1	2
	Total	25	1.60	.707	.141	1.31	1.89	0	2

Table 35: Statistics for Q4a – Q7a and Overall scores.

The descriptive statistics for the subjects that were asked these supplementary questions are given in Table 35 and the percentages for each question are given in Figure 89. The Kolmogorov-Smirnov test for normality reveals that the score distributions for Question 4a, Question 6a, and Question 7a are non-normal ($p = 2.3 \times 10^{-5}$, $p = 1.36 \times 10^{-5}$, and $p = 0.002$ respectively) while the Overall and Question 5a distributions are approximately normal ($p = 0.379$ and $p = 0.167$ respectively). Note that by Overall scores in this instance, we mean overall scores on Question 2 through Question 7 and Question 4a through Question 7a.

The non-parametric Kruskal-Wallis test (for non-normal distributions) suggests that there is a moderate statistical effect ($p = 0.063$) for Question 4a but the scores on Question 6a and Question 7a are not statistically significant ($p = 0.968$ and $p = 0.881$). A two-way ANOVA on the Overall and Question 5a scores both show a moderately significant main effect on facial feature visibility ($p = 0.059$ and $p = 0.081$ respectively). There are no statistically significant gesture visibility or interaction effects. The estimated marginal means of the Overall score and the score on Question 5a are shown in Figure 87 and Figure 88. Again, none of these results are significant at the $\alpha = 0.05$ level, but they represent a moderately significant effect at an $\alpha = 0.1$ level.

Note that if we consider the Overall score of Question 4a through Question 7a, we get a very similar result, with the two-way ANOVA giving a moderately significant effect for head visibility ($p = 0.053$). Therefore, the effect we see on overall score is consistent when considering the overall score of all 11 questions (Q2 – Q7 and Q4a – Q7a) and the overall score on the four supplementary questions (Q4a – Q7a).

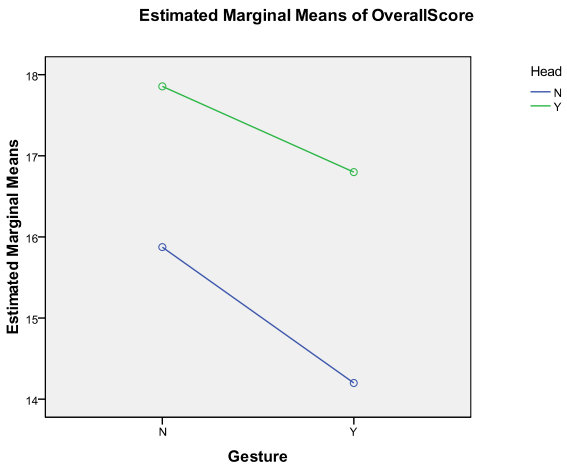


Figure 87: Means for the Overall scores on the extended questionnaire.

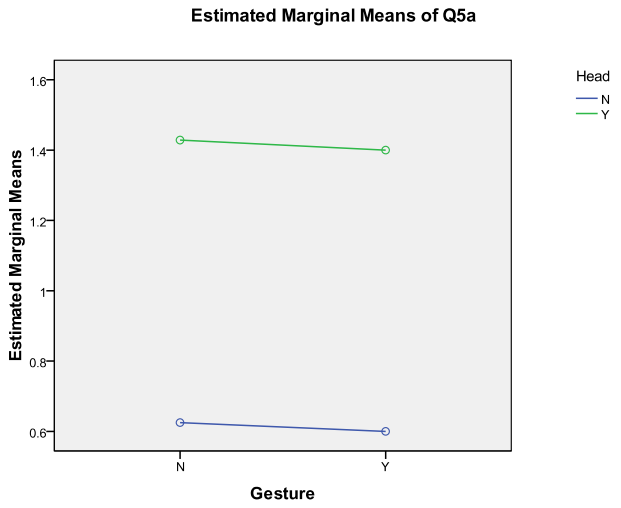


Figure 88: Means for Question 5a scores on the extended questionnaire.

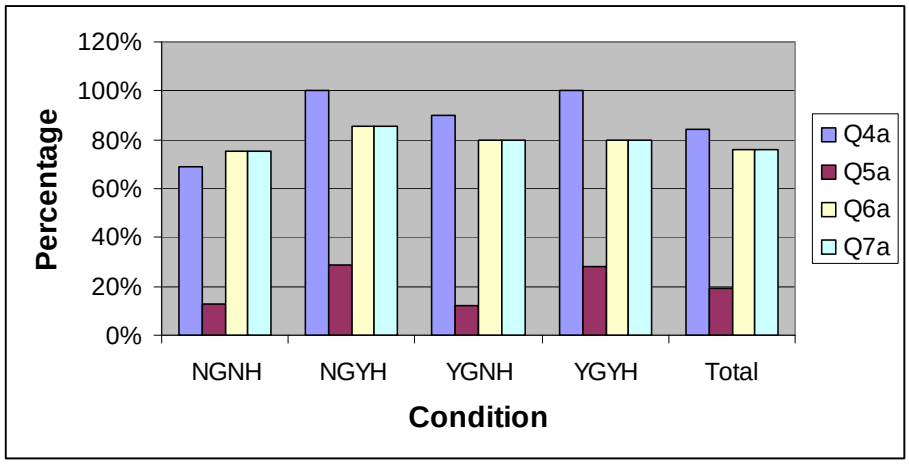


Figure 89: Percentage scores per question across conditions

15.14.2.4 Question 5a in Detail

One interesting result from this analysis is the relatively poor scores on Question 5a. Recall that the maximum score for Question 4a, 5a, 6a, and 7a are two, five, two, and two respectively. Question 5a is marked out of five because the subject is asked to list all of the "global disasters" that are presented in the lower right quadrant of the Pascal's Wager diagram (there are five). The average scores for the four questions are 84%, 20%, 76%, and 76% respectively (see Figure 89).

In order to understand the poor scores recorded by the subjects on Question 5a we consider the fixation times in the acts in which the lower right quadrant has artifact AOIs.

The “disaster” artifacts that provide the answer to Question 5a are created by the presenter revealing the list through a sharp action. In the YGYH condition, this is done by ripping a paper off of the whiteboard, revealing the list of five disasters underneath. In the three other conditions, the list “appears” on the whiteboard at the same time the paper manipulation occurs in the YGYH condition.

There are twelve scenes in Act 4 where the presenter refers to the “disaster” quadrant using artifact gestures and artifact manipulation. Four of these scenes are before the list is revealed, pointing to the bottom right quadrant when no list is visible (two explicit pointing and two explicit pointing post scenes). Two of the scenes (Scene 4-36 and 4-37) are the artifact creation scenes. The first is a relatively short artifact manipulation scene where the artifact is revealed and the second is a post artifact manipulation scene that immediately follows the artifact creation scene. We consider the ArtifactManipPost scene part of the creation process because the creation scene is very short (500 ms) and the subject’s gaze tends to linger on the list after the action takes place.

There are six scenes in the remainder of Act 4 in which the presenter refers to either the list or the quadrant in which the list exists. Gaze fixations from these pointing gestures are captured using our traditional ImplicitPointArtifact (four scenes) and ImplicitPointArtifactPost (two scenes) AOIs. We break the scenes down as follows for analysis. First, we consider only the artifact creation scenes (Scenes 4-36 and 4-37), analyzing the ArtifactManip and ArtifactManipPost AOIs. We then consider all of the scenes involving the bottom right quadrant of the diagram. We consider all of the artifact related AOIs (ArtifactManip, ArtifactManipPost, ExplicitPointArtifact, ExplicitPointArtifactPost, ImplicitPointArtifact, and ImplicitPointArtifactPost) across these scenes as a group. We also consider the FacialFeature AOIs across all twelve of these scenes. Lastly, we consider the total fixation time in all AOIs across all of the scenes.

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
DisasterArtifactAOI	YGYH	10	2000.70	476.536	150.694	1659.81	2341.59	1000	2393
	NGNH	9	1500.89	625.686	208.562	1019.94	1981.83	399	2374
	YGNH	10	1605.70	553.990	175.187	1209.40	2002.00	300	2294
	NGYH	8	1697.13	389.567	137.733	1371.44	2022.81	1016	2133
	Total	37	1706.73	536.150	88.143	1527.97	1885.49	300	2393
DisasterGestureAOI	YGYH	10	5851.80	1633.113	516.436	4683.54	7020.06	2480	8793
	NGNH	9	4437.44	1961.419	653.806	2929.76	5945.12	1458	8236
	YGNH	10	4205.30	1754.977	554.972	2949.87	5460.73	1299	7498
	NGYH	8	3719.00	1122.948	397.022	2780.19	4657.81	1734	4884
	Total	37	4601.62	1787.319	293.833	4005.70	5197.54	1299	8793
DisasterFaceAOI	YGYH	10	2479.30	1312.994	415.205	1540.04	3418.56	219	4409
	NGNH	9	39.89	79.152	26.384	-20.95	100.73	0	180
	YGNH	10	4.00	12.649	4.000	-5.05	13.05	0	40
	NGYH	8	3297.75	1131.852	400.170	2351.50	4244.00	1058	4725
	Total	37	1393.89	1676.067	275.544	835.06	1952.72	0	4725
DisasterTotalAOI	YGYH	10	8331.10	2322.603	734.472	6669.61	9992.59	3540	11564
	NGNH	9	4477.33	1946.371	648.790	2981.22	5973.45	1458	8236
	YGNH	10	4209.30	1756.948	555.596	2952.46	5466.14	1299	7498
	NGYH	8	7016.75	2173.926	768.599	5199.30	8834.20	2792	9609
	Total	37	5995.51	2662.826	437.766	5107.68	6883.34	1299	11564

Table 36: Descriptive statistics for the Disaster scenes.

The distributions of all of the artifact fixation measures are approximately normally distributed (Kolmogorov-Smirnov test for normality) with the exception of the FacialFeature AOI fixations. The non-parametric Kruskal-Wallis test on the FacialFeature AOI fixation times indicates that the mean ranks are significantly different ($p = 1.3 \times 10^{-6}$). This can be seen clearly from the FacialFeature descriptive statistics in Table 36, as the amount of FacialFeature fixation time when facial features are not visible is negligible.

A two-way ANOVA to test the mean fixation time for the artifact creation scenes (the ArtifactManip and ArtifactManipPost AOIs) gives a moderately significant effect for facial feature visibility ($p = 0.096$). As in the previous analysis, it appears that facial feature visibility can increase the artifact fixation time by helping to focus attention in the

right area. There is no significant main effect for gesture visibility (p = 0.245) and there is no interaction effect (p = 0.568).

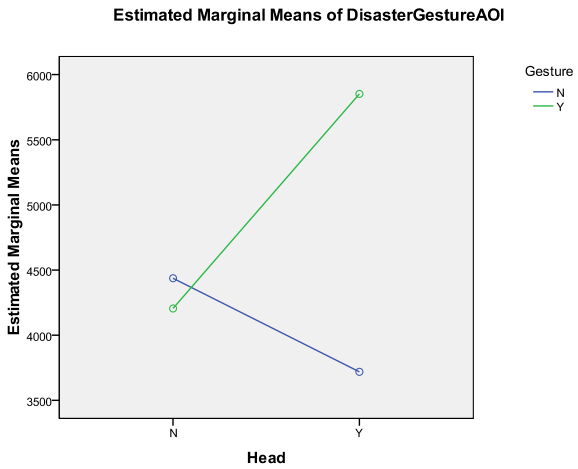


Figure 90: Estimated means for artifact AOIs across the disaster scenes.

A two-way ANOVA to test for differences in the mean fixation time for all artifact AOIs across all 12 scenes reveals a significant interaction effect (p = 0.039) between facial feature and gesture visibility. When facial features are visible there is a significant difference across the gesture visibility conditions, with significantly more artifact fixation when gestures are visible (Tukey HSD, p = 0.05). When facial features are not visible, gesture visibility has very little impact on artifact fixations (Tukey HSD, p = 0.990). This can be seen pictorially in Figure 90. This implies that attracting attention to artifact AOIs in these scenes is done most effectively through having both facial feature and gesture visible.

A two-way ANOVA to test for effects on the mean overall fixation time across all AOIs and all twelve scenes reveals a significant main effect for head visibility ($p = 2.46 \times 10^{-5}$). There are no main effects for gesture visibility (p = 0.447) and there are no interaction effects (p = 0.253).

Finally, we consider the overall percentage of fixation time spent in both the artifact based AOIs and all AOIs. Subjects spent on average 27% of the total scene time in artifact AOIs and 36% of the total scene time in some type of AOI. If these percentages are considered in the context of the fixation time percentages for different gesture types

shown in Figure 59, these percentages are certainly within the bounds of typical fixation percentages for ExplicitPointArtifact and ImplicitPointArtifact AOIs.

Based on this analysis, it is difficult to explain why Question 5a was so poorly answered. Overall fixation percentages for both artifact AOIs and overall AOI fixation time are within the bounds of typical fixation times for scenes with those AOI types. Thus we can not claim that subjects attended to the artifact AOIs less in the disaster scenes than in other scenes. The artifact creation scenes, although not as effective as writing scenes, were quite effective at drawing attention to the list of disasters when the artifact was created. The one main effect that is not present as strongly as in other scenes is the fact that there is no main effect for gesture in any of the artifact creation scenes (Scene 4-36 and 4-37), the scenes in which the disaster artifacts are referred to, or the total AOI fixation time across all scenes. Despite the fact that this main effect is not present, there is a strong interaction effect in the disaster artifact AOI related scenes in which gesture visibility has a strong positive effect on the mean AOI fixation time.

15.14.2.5 Task measure summary

Although the analysis of the subject's questionnaire responses are not statistically significant at the $\alpha = 0.05$ level, there are some moderately significant effects across some of the conditions. In particular, in the expanded questionnaire responses the Overall score and Question 5a score both show a moderately ($p < 0.1$) significant effect on head visibility. We discuss these results, in particular in the context of the gaze fixation data, in our summary analysis in Chapter 11.