Adaptive Mediation in Groupware

A Dissertation

Presented to

The Faculty of the Graduate School of Arts and Sciences

Brandeis University

Department of Computer Science

Professor Timothy J. Hickey, Advisor

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by

Joshua E. Introne

August 2008

The signed version of this signature page is on file at the Graduate School of Arts and Sciences at Brandeis University.

This dissertation, directed and approved by Joshua E. Introne's Committee, has been accepted and approved by the Faculty of Brandeis University in partial fulfillment of the requirements for the degree of:

DOCTOR OF PHILOSOPHY

Adam B. Jaffe, Dean of Arts and Sciences

Dissertation Committee:

Prof. Timothy J. Hickey, Computer Science

Prof. James Pustejovsky, Computer Science

Prof. David Jacobson, Anthropology

Dr. Brant A. Cheikes, MITRE Corporation

Dr. Bradley A. Goodman, MITRE Corporation

ACKNOWLEDGEMENTS

My journey down the long road to the completion of this Ph.D has not always been a smooth one. But one of the few regularities that has marked my journey has been appearance of individuals who helped me take the next step and to grow beyond ephemeral limitations to reach just a little further.

I am grateful to Professor Alterman, my first advisor and the individual perhaps most responsible for the body of knowledge I have accumulated over the years. He pointed me to a research direction when I really only had the vaguest of ideas about what I sought to accomplish, and got me started on my path.

In those initial years working in the GROUP lab, Dr. Alex Feinman and Dr. Seth Landsman were commiserators, collaborators, teachers, and friends. I am indebted to them both for their work on VesselWorld, and our collaborative efforts taught me volumes about how to code, and how to be part of a team. Throughout it all, their constant and ready advice helped me to learn how to cope with the trials of graduate school.

As the financial burdens of graduate life began to weigh more heavily, I was lucky to find employment at Charles River Analytics, Inc., and especially lucky to find Dr. Subrata Das, without whom the last half of this thesis could not have been written. Dr. Das has been a constant and enthusiastic supporter and friend, and his work and research are the

basis of the REASON platform. I'm ever grateful to Dr. Das for being a tremendous mentor, and for being so adept at building my self-confidence while challenging me to think rigorously and to always exceed my prior accomplishments.

There are many others at Charles River who contributed directly and indirectly to this dissertation. I am very grateful to Dr. Greg Zacharias, who was always very supportive of the REASON effort, and graciously allowed me to use the code developed by myself and others to execute the second case study in this thesis. Dr. John Langton, Andrew Young, Scott Harrison, and many other developers contributed much to the design and development of the platform.

The REASON effort would not have been possible without the support and confidence of Chris Ramming at DARPA. His enthusiasm and vision are what made the platform something more than just a nice idea. The trajectory of this project arcs long and further than any of us may have realized, and this is a direct result of the force Chris applied at its outset. I am personally grateful for the support Chris has provided to me directly throughout the years. The confidence of one for whom you have enormous respect is a humbling and heady surprise.

There are many others who contributed in large ways to the REASON project. Dr. Norman Warner provided the experimental materials that formed the basis of my case study, and Johann Larusson was always there to offer technical support and lend a ready ear and words of encouragement. And there are too many friends and family to mention here who were willing to spend their time participating in my pilot study, and many others who took their time to help me figure out how ultimately the experiment should be

conducted.

The dissertation construction process was shepherded by a fantastic committee. I'm grateful to Prof. James Pustejovsky and Prof. David Jacobson for their continued encouragement and support, and likewise to Dr. Brad Goodman for his careful reading and welcome advice. I am especially indebted to Dr. Brant Cheikes for his numerous pointers to others' research, and careful and insightful analysis of mine. As a result of his advice, many avenues for future work have become clear, and I am eager to pursue them.

I reserve special thanks for Prof. Tim Hickey, who willingly shouldered the burden to become my last minute advisor. Tim has provided constant and steady support, a critical eye, and invaluable feedback. Working with Tim has been a fantastic experience, and my only regret is that I've had such a small amount of time to learn from him all that I have.

Outside of direct involvement with my dissertation work, I have had many amazing friends who have have always been there to prop me up and help me to relax a little. I cannot thank everyone properly here. I am indebted to Bill and Amy for always welcoming my wife and I when the discordance of urban living seemed too much to bear, and for filling our lives with love and laughter. And to Drs. Long and Gray for the continual and aggressive embrace of all that is wonderful and is right in front of us, like Ardbeg Uigedail, and an occasional popsicle. And to Sachin & Sara, and Courtney & Maura, and Tom & Marissa & Tess, and Dan & Laura, Duane & Laura, Mark & Betsy, Jay & Jenny, Kevin & Kristina, Martin & Leah, Billy, Vitaly & Abby, Russell & Jenny, Brandon, and many, many others.

I cannot thank Mom & Dad enough. They are at the core of everything positive in my

life, and I have been so lucky to be blessed with such wonderful parents. They are some of my best friends and greatest role models, and their unwavering support sustains me always.

My other family has only added to this support. Nana, Lucas and Minnie, and my more recently acquired family – Karen (Mom') and Mary Lou (Mom''), Mike & Missy, Bella & Jeff, and Ben & Rachel & Trajan – have not only displayed enormous patience for my constant absence from family functions, but have continuously offered their love and assistance in spite of it all.

Finally, my beautiful and brilliant wife Fredrica has been my best friend, my muse, my unflagging copy-editor, cheerleader, and true love. All at the same time, and without ever really breaking a sweat. I may have written a dissertation, but Fredrica deserves at least as much credit as I do for finishing my Ph.D.

Thank you all.

ABSTRACT

Adaptive Mediation in Groupware

A dissertation presented to the Faculty of the
Graduate School of Arts and Sciences of Brandeis
University, Waltham, Massachusetts

by Joshua E. Introne

Groupware is software that is designed to support groups of people engaged in a task. As with single-user software, groupware mediates the users' interaction with a domain in pursuit of a goal, but when using groupware, the system also mediates the interactions users have with one another. In this dissertation, I will show that this aspect of groupware presents software developers with two opportunities that may be exploited when developing adaptive systems. The first opportunity concerns the knowledge acquisition problem. This opportunity presents itself because information that collaborators need to exchange so they can stay coordinated in their shared task will pass through the system itself. I will show that it is indeed possible to use this information, and that there is a repeatable technique for doing this that can be applied in other systems.

The second opportunity concerns how adaptive techniques may be applied once this

information is attained. Because the system itself mediates all of the collaborators' interactions, the developer has the ability to “reach beyond” the software itself, and transform the collaborative process into one that is closer to a notional ideal. I call this adaptive mediation, and this approach represents a relatively novel application of adaptive technology. I will show that it is possible to use knowledge obtained by exploiting the first opportunity in order to transform a group's collaborative process as described.

These two opportunities are the focus of this dissertation, and my goal is to show that they exist, and that it is possible to take advantage of them. To do this, I present two case-studies. The first of these illustrates a solution to the knowledge acquisition problem in an adaptive groupware system. The second incorporates this solution, and uses the acquired knowledge to address a well-known problem that afflicts group information processing. Both case-studies successfully exploit the identified opportunities, and the second case study leads to some new techniques for and insights about group decision support.

Table of Contents

Introduction

Groupware is software that is designed to support groups of people engaged in a task (Ellis, Gibbs, and Rein 1991). As with single-user software, groupware mediates the users' interaction with a domain in pursuit of a goal, but when using groupware, the system also mediates the interactions users have with one another. In this dissertation, I will show that this aspect of groupware presents software developers with two opportunities that may be exploited when developing adaptive groupware. The first opportunity concerns the *knowledge acquisition problem.* This opportunity presents itself because information that collaborators need to exchange so they can stay coordinated in their shared task will pass through the system itself. I will show that it is indeed possible to use this information, and that there is a repeatable technique for doing this that can be applied in other systems.

The second opportunity concerns how adaptive techniques may be applied once this information is attained. Because the system itself mediates all of the collaborators' interactions, the developer has the ability to “reach beyond” the software itself, and transform the collaborative process into one that is closer to a notional ideal. I call this *adaptive mediation,* and this approach represents a relatively novel application of adaptive technology. I will show that it is possible to use knowledge obtained by exploiting the first opportunity in order to transform a group's collaborative process as described.

These two opportunities are the focus of this dissertation, and my goal is to show that they exist, and that it is possible to take advantage of them. To do this, I present two case-studies. The first of these illustrates a solution to the knowledge acquisition problem in an adaptive groupware system. The second incorporates this solution, and uses the acquired knowledge to address a well-known problem that afflicts collaborative decision making. Both case-studies successfully exploit the identified opportunities.

The second case study also leads to several contributions to the field of group decision support. The group decision support system developed for the study is a novel approach to decision support. It combines argument visualization with more traditional decision analytic techniques, and uses the interlocutors' deliberative conversation as an analytic decision model. Furthermore, an analysis of the way the system was used offers some new insights about and more detailed understanding of collaborative story formation in some kinds of decision problems.

The remainder of this introduction will contextualize these contributions, elaborate upon the findings presented in this dissertation, and provide the reader with a roadmap for the following chapters.

1 Adaptive Systems

In order to understand why the opportunities discussed above exist, and why they are important, it will be necessary to first say a little about what adaptive systems are, and why they are used. Unfortunately, defining adaptive software as a class is not as simple as defining groupware. Several definitions have been offered (e.g. Benyon 1993; Langley 1999; Oppermann 1994), but these tend to focus upon aspects of adaptivity, and exclude many systems that appear in the literature on adaptive systems. For purposes of this discussion, I begin with the following working definition:

> *Adaptive software is software that uses some runtime information about the domain, the user, or the system itself to improve the user's domain activity with respect to an ideal.*

What constitutes an "improvement" in the user's domain activity varies between systems, but it always implies an ideal or theory about that activity that was assumed by the designer. In the vast majority of cases, this ideal has to do with using the system itself. That is, there is a presumed "ideal" interaction with a given software system, and the adaptive technology is incorporated to get the user a little closer to it.

The reason adaptive technology is used in this manner is because it is hard to design software that is both flexible (i.e. can be used for many tasks by people with different needs and abilities) and meets this ideal. The simple reason for this is that it is possible for a user to accomplish more things with a small set of low-level operators in a given domain than with an equal number set of higher-level operators in the same domain. This is why I type on a keyboard composed of letters, and not words.

Thus, there are a spectrum of possible designs for any given piece of software bracketed by two notional design strategies each with its own limitations. At one end of this spectrum are designs that seek to identify the "lowest common denominator" set of low-level operators that will allow users to construct all of the possible high level operations they might need (e.g., a shell scripting language at the UNIX command line). At the other end of the spectrum are systems that attempt to identify the union of all possible high-level tasks all users might want, and pack them into the interface. The former has the drawback of requiring the user to string together more operations, and to remember more such sequences. The latter makes it difficult for the user to figure out which operator they need among a proliferation of additional operators that may not be relevant to their task.

Adaptive software can help out at either end of the spectrum. To contend with the problems on the "low-level operator" side, adaptive techniques may be used to recognize what the user's goals are, and subsequently automate sequences of operations for the user. To contend with the problem of too many high-level operators, an adaptive system might recognize something about the user that indicates which kind of tasks they might want to do, and customize the interface to present just those tasks.

This brief introduction to adaptive software is a vast over simplification. Design does not actually take place at either end of this notional spectrum, and there are other approaches to adaptive systems aside from those presented in this introduction. I will discuss some of these and offer more support for this analysis in Chapter 2. Nonetheless, the preceding framework does characterize a good portion of mainstream research in adaptive systems, and provides a backdrop for the following discussion of difficulties with and limitations of existing adaptive systems research.

1.1 The Knowledge Acquisition Problem

The development of adaptive systems is complicated for many reasons, but a fundamental limitation on the scope of adaptation is the amount and type of information the system can obtain at runtime (see Terveen 1995). For instance, consider the case where the system attempts to infer the user's runtime goal in order to automate some of the user's interaction with the system. If the interface consists of low-level operators, the system's job is to match the sequence of operators to a plan library of known tasks.

Unfortunately, this plan-recognition problem is widely acknowledged to be very hard for realistic domains. Constructing a complete and correct plan library can be a daunting task (see Lesh et al. 1996). Furthermore, the user may or may not be performing the operational sequence correctly, tasks may overlap, and the system has no way of knowing when a task begins and when it ends. This is the “keyhole plan recognition” (P. R. Cohen, Perrault, and J. F. Allen 1981) problem, and the solution is to find a way to get a little more information from the user to help the system make better guesses about the user's intentions.

One way to do this is to employ a “conversational agent” at the interface that “collaborates” with the user about the user's task. For example, whenever I type “Dear” at the beginning of a document, a conversational agent could leap into action and ask me if I'd like to write a letter. Unfortunately, this style of interaction is often perceived as disruptive by the user.

Another possibility might be to develop interfaces that somehow organize the operations the user might like to make in a high-level task-relevant manner (see St. Amant, Dinardo, and Buckner 2003). In this manner, whenever the user takes an action in the interface, the system has a better idea about what she is trying to accomplish, and plan recognition is simplified. This approach is often very successful, and leads to powerful human collaborative systems. However, it does not solve the original design problem. If indeed the domain can be constrained sufficiently to build an interface around high-level tasks instead of low-level operators, there may be no design problem to be solved.

The above paragraphs illustrate the knowledge acquisition problem for adaptive systems. That is, how do we get the information we need from the user, without disrupting the user, in those systems where that information is not readily available? The first contribution of this dissertation is to present a potential solution to this problem in groupware systems.

1.2 The Boundaries of System-Oriented Adaptation

My second observation is not about the difficulty in implementing adaptive systems, but rather about the difficulty the field has had in moving beyond improving less than optimal design.

Some researchers have criticized adaptation as being a "band-aid" solution to a problem that deserves better design techniques (e.g. Maes and Schneiderman 1997). There are a variety of responses to this accusation. One response is that computers offer a set of "human complementary" abilities, and that adaptation offers a way to bring these abilities into a productive human-computer collaborative partnership (Terveen 1995). A similar observation is that computers are well-suited to serve as proxies for people in a variety of relatively mundane tasks, freeing up time and resources that could be devoted to those tasks requiring human level intelligence (Maes 1994).

Both of the above counter-arguments suggest that it is possible for adaptive technology to move beyond the boundaries of the system itself, and there are certain classes of adaptive systems that do this successfully. Intelligent tutorial systems can serve as a proxy for a teacher, and further the domain goal of educating a student. Information retrieval and recommender systems allow people to find things in much larger collections of information than would otherwise be possible. Augmented reality systems offer the potential of extending our perceptual abilities beyond what is possible with our unadorned biological hardware. Safety critical systems help us to avoid disastrous errors in dangerous domains (and the adaptive spelling technology incorporated into my word-processor helps me spell words like disastrous).

Nonetheless, the vast majority of adaptive systems are designed to improve a user's engagement with a piece of software, and there are only a few classes of adaptive technology that seek to improve the user's domain activity directly. It is hard to find new application domains for adaptive systems that are both within our capabilities to develop and leverage computer technology to extend our abilities as actors in the world.

The second contribution of this thesis is to identify a class of adaptive system that assists groups of people with more than just use of the system itself but has not yet been broadly explored. I will do this by describing a successful case study with a system that is a member of this class.

2 Solutions in Mediated Collaboration

This introduction opened with the observation that two opportunities for adaptive systems developers present themselves because groupware mediates communication between collaborators engaged in a common task. With the background provided in the previous two sections, it is now possible to say a little more about precisely how mediation in groupware creates these opportunities.

2.1 Solving the Knowledge Acquisition Problem

The knowledge acquisition problem in groupware may potentially be solved because some portion of the information that collaborators pass through the system supports coordination in their domain task. Consequently such inter-user communication can be valuable resource for a system trying to infer something about their domain task at runtime. The trick, of course, is in how to get at this information.

The solution lies in the fact that this coordinating information frequently has a common underlying structure, and it makes up a good portion of the collaborative dialog. Many other researchers have pointed out that for coordinating information that has these properties, it is not uncommon for people to engineer representational artifacts that can simplify its exchange, serve as reminder to the users about what the important pieces are, and make it available for subsequent use (e.g. Suchman and Trigg 1992; Hutchins 1995; C. Goodwin and M. H. Goodwin 1996; Schmidt and Simone 1996). People use these representations because they want to use them, and because they simplify coordination in the domain task. These representations do not interfere with the domain task; they are integrated into it and improve it.

The structure that this kind of mediating representation lends to information also makes the information easier for the computer to process. Hence, the insight that will be elaborated in subsequent chapters is that solving the knowledge acquisition problem in groupware amounts to identifying or creating a structured, mediating representation that improves coordination in the collaborators' shared task (see Alterman et al. 2001, Introne and Alterman 2006) .

2.2 Extending the Scope of Adaptation

Because the system mediates collaboration, it is possible to influence the collaborative process. All that is required (other than the observation that this is possible) is a vision for an ideal collaborative process that the developer can use in constructing the adaptive functionality.

There are few examples of how this may be done. One of the few can be found in the PIERCE / Epsilon system proposed by Goodman et al. (2005), and based upon the work of (Soller 2004). PIERCE is an agent designed to integrate with Epsilon, a collaborative learning environment and intelligent tutorial system. PIERCE is designed to monitor the group's collaborative activities for dysfunction, and “jump in” when it notices these in order to guide the group back towards the collaborative ideal designed into the adaptive system.

In order to solve the knowledge acquisition problem, Epsilon requires the collaborators to tag their dialog with speech acts (Searle 1969). This, unfortunately, is not a mediating structure that serves any purpose for the collaborators. Nonetheless, PIERCE embodies a novel adaptive strategy. Instead of adapting the users' interactions with the system to improve use of the system itself, the system mediates interactions between the users in order to achieve an ideal in the domain of collaborative learning.

I will illustrate an adaptive groupware platform that is designed to help people overcome biased collaborative information processing within the domain of group decision making. The system leverages mediating structure that is incorporated for the purpose of helping users to better coordinate their deliberative process. The system illustrates both a novel solution to the problem of biased collaborative information processing, and a novel adaptive strategy in groupware.

3 Contributions to Group Decision Support

In addition to the contributions discussed above, the implementation of the adaptive group decision support system discussed above, and the subsequent analysis of its use, contribute to the field of group decision support in several ways.

First, the implemented system is a realistic strategy for group decision making. The system was generally well received, and collaborators were able to use its incorporated functionality to good effect with minimal training. In comparison to other group decision support systems it is unique because it enables people to have a relatively normal discussion about a decision problem while constructing an analytical model of that decision. This is enabled by the use of argument visualization as a mediating representation, which addresses the knowledge acquisition problem without disrupting the deliberative process. The analytical model is in turn used to guide people to a collaborative decision that is more consistent with their exchanged information.

In analyzing data from this case study, it was observed that people without the adaptive system (henceforth the "non-adaptive" groups) made decisions in an entirely different fashion than those with the adaptive system. Although the non-adaptive groups did not make decisions that were consistent with the balance of exchanged information, they did make the "correct" decision just as frequently as the groups with the adaptive system.

The critical difference between the two conditions studied was that people using the non-adaptive system engaged in a more collaborative story construction process en route to solving the decision problem. These qualitative differences were persistent enough to be isolated via a quantitative metric that identified regions of the conversation where people had highly focused, consensual discussion. Additional analyses illustrated that while both groups engaged in a story construction process, groups without the adaptive system created shared stories that contained information and inferences from more members of the group than in the adaptive case.

Reflecting upon the case study suggests that one of the reasons these observations were possible was because of the research design and approach to adaptation that were employed. Two conditions were examined in the study, and in one of the conditions, an idealized model of decision making was "thrust" upon collaborative groups. This model was based upon well-supported mathematical models of decision making in the decision analysis literature. Because all group communication was captured in these studies, it was possible to examine the data to see precisely how this idealized model works, and does not work, with the "natural" collaborative process. This research design may serve as a template for future studies that will allow us to understand collaborative processes in general, and test our theories about how collaboration should occur.

4 Roadmap

This dissertation is organized as follows. In the following chapter, I will present an overview of various kinds of adaptive systems. As with many reviews, I do not intended to exhaustively cover the field. Instead, I will seek to elucidate the historical roots of adaptive systems and offer a broad overview of the kinds of things that adaptive strategies have been used for. This review is organized according to a taxonomy which is a useful way to characterize the field. Throughout the review, I will identify the techniques that have been used to acquire the user knowledge necessary to drive the adaptive functionality. At the end of the review, I will summarize the knowledge acquisition problem for adaptive systems and explain what I consider to be limitations in the scope of existing adaptive systems research. I will conclude with an explanation of how my work seeks to address both of these concerns.

In the third chapter, I present my first case study. The study was run using a platform called "VesselWorld," which is a notional C2 (command and control) application designed to mediate a search and rescue operation. In the case of VesselWorld, search and rescue is performed in a harbor, and the objects to be "rescued" are toxic waste barrels. The platform is designed for three collaborators, and is turn based. I will show how the introduction of a mediating artifact to keep track of the toxic waste barrels produces information that can support adaptive functionality that improves use of the system, and also eliminates some domain errors. I will offer direct support for the utility of the structured information via an analysis of the inference engine's performance with collected data. I will also analyze the use of an interaction-oriented adaptive component that was implemented on top of this inference engine.

The fourth chapter will present a review of literature on decision making. The review will cover models for, approaches to, and analyses of individual decision making. I will also describe problems that are unique to collaborative decision making, and tools that have been developed to support it. There is a very large body of literature on these topics, and I will focus upon just those elements that are relevant to framing the case study in Chapter 6. I will pay particular attention to Pennington and Hastie's (1992) examination of story based decision making in juries. Their work explains to a large degree the phenomena I witnessed in my own research.

The fifth chapter will briefly pull-together research reviewed in the preceding chapter to arrive at a design of the adaptive group decision support system that was the platform for my case study. This chapter also covers some implementation details, and provides screenshots of the implemented application.

The sixth and seventh chapters describe a case study with the group decision support platform, and present in-depth analyses of the collected data. The case study itself is based upon materials developed by Stasser and Stewart (1992) for a "hidden profile" experiment that was designed to investigate a particular information processing asymmetry in group decision-making. As I will show in Chapter 6, the adaptive functionality included in the platform was successful in overcoming the previously identified asymmetry. However, the adaptive functionality did not reliably improve the ability of collaborators to find the "correct" solution beyond the ability of those without the adaptive functionality. This observation forms the basis for an extended analysis, which is presented in Chapter 7. In this second analysis, I will show that collaborators without the adaptive functionality were better able to construct collaborative stories. Evidence of this is presented via the introduction of a quantitative metric that is able to isolate regions of critical story construction for the group. Because these groups were effective collaborators, they were able to discover the correct answer, even when their exchanged information didn't seem to warrant such an answer.

Finally, in Chapter 8, I will summarize my work, and catalog the numerous avenues for future work that have come about as a result of my research. A substantial portion of this chapter will be devoted to expanding upon, or addressing gaps in the presented work, that should be addressed so as to complete my analyses and strengthen my conclusions. An equally large portion will be devoted to further, future research. In particular, I will focus upon extensions to the implemented decision support platform to address user needs discovered during the experiment. I will also propose new mediating structures for group decision-making that are supported by the analysis presented in Chapter 7. I will conclude with a brief discussion about value of the experimental design employed in Chapters 6 and 7.

Chapter 1: Approaches to Adaptive Systems

Adaptive techniques are very often used to address certain kinds of design problems, and so to understand why adaptive techniques are employed, it is helpful to understand why these design problems occur in the first place. So, as a preface to my analysis of adaptive software, I offer the following quick analysis of one class of design problems that was instrumental in the birth of adaptive software.

Software systems are tools, usually designed to help the user accomplish some goal in a domain. A software system can be viewed as an artifact that both represents the domain and mediates a user's interaction with that domain (Norman 1991). It presents the user with a set of controls and feedback channels that allow the user to interact with the represented domain. Because software mediates this interaction, it can also change its nature.

Generally, the goal of this mediation is to transform the user's interaction with a domain so that it becomes easier for the user to accomplish her goals. In an ideal system, the user would simply express her intention in some very direct manner, and it would be done. This is the tool builder's ideal, related to Heidegger's notion of *ready-to-hand*, which describes how the tool can become an extension of the actor, essentially disappearing (Heidegger 1962). Explicit references to Heidegger are made by others (Bødker 1990), and the idea is elaborated upon in many writings on design and usability. Norman (1991) speaks of the designer's task as minimizing the "gulfs" of evaluation and execution for the user. Hutchins, Hollan, and Norman (1985) speak of "semantic" and "articulatory" directness in software; the former is concerned with whether or not and how concisely the user can express her intentions to the system; the latter is concerned with the relationship between symbolic expression in the interface and the meaning of that expression. Even Shackel's (1991; pg. 24) definition for usability - "the capability to be used by humans easily and effectively" - is related to this ideal.

For purposes of conciseness and ease of reference, I will refer to this idea within the context of software as the "software interaction ideal," defined (quite loosely) as:

> **The Software Interaction Ideal**: *The user expresses to the software as directly as possible what they wish to accomplish in the domain, and the software just does it!*

One established way to achieve the software interaction ideal is through the use of direct manipulation interfaces (Shneiderman 1983). Direct manipulation is a powerful technique that can be used when it is possible to identify "objects" in the domain that are the focus of the user's intentions. Moreover, it must also be possible identify a set of manipulations with those objects that map closely to the user's domain intentions (Hutchins, Hollan, and D. A. Norman 1985; D. A. Norman 1991). For instance, to move a file on my computer, I can simply drag it from one "folder" to another. The file is an "object," and a "drag operation" expresses my intention to move the file in the file system.

Direct manipulation is not always feasible, nor is it always the best approach. Domain intentions may exhibit enough variance that it makes more sense to provide the user with a set of lower level operators, perhaps in the form of a command language, or menu driven system. The inclusion of lower level operators at the interface means that a given piece of software can be used to do a broader range of things in a particular domain. But the inclusion of these operators also increases software complexity, making it harder to understand. It also requires that the user do more work to achieve any given goal in the represented domain. This stands in direct opposition to the software interaction ideal stated above, and is a fundamental tension in design.

In the simplest cases, adaptive techniques are a means by which to address the fundamental tension between software flexibility, and the software interaction ideal. This is the historical basis for adaptive software systems, and the starting point for my analysis.

In the following sections I will cover several aspects of adaptive software. I will offer a brief historical perspective regarding the birth of adaptive systems to lend support to the above assertion. Following that analysis, I will offer a two-dimensional taxonomy that can be used characterize the domain of adaptive systems, and describe a number of examples to help fill out the taxonomy

After reviewing a broad array of systems, I will discuss two limitations in adaptive systems research. The first is the knowledge acquisition problem, which is commonly acknowledged to be a difficult hurdle in implementing adaptive systems. The second limitation concerns the breadth of the field. I will then describe the opportunities to overcome these limitations in groupware systems. The remainder of this dissertation will be focused on elaborating those opportunities.

1 The Roots of Adaptive Systems

During the mid-eighties, computer-scientists investigated techniques for simplifying the construction of user interfaces according to specification. One approach that was developed was the user interface management system (UIMS).

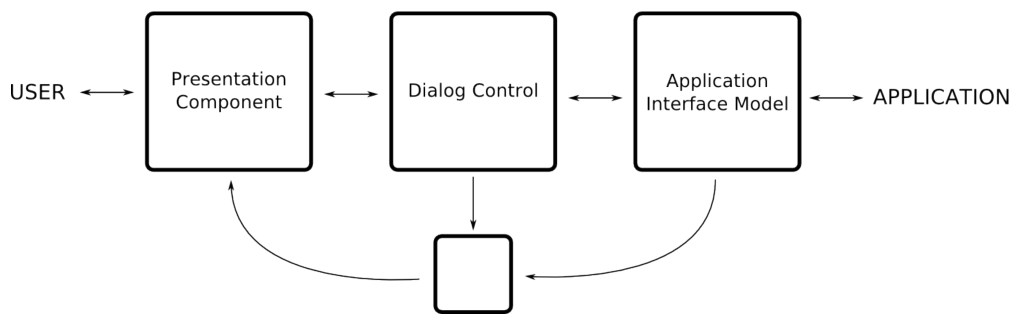


Figure 1: The Seeheim Model

One influential model in the development of UIMSs was the Seeheim model of user interfaces (Green 1986; Pfaff 1985). The Seeheim model divides the user interface into three main components (Figure 1). The intention of this model was to separate the domain representation (application interface model) from the interface representation (presentation component), hence allowing the interface to change independently from the underlying system.

Although the Seeheim model turned out to have some practical difficulties, it is illustrative of the kind of modular design ideas that were being considered at the time. In this regard, the Seeheim model was very similar to other software engineering approaches begin explored at the time, such as the Model, View, Controller pattern introduced with Smalltalk-80 (Krasner and Pope 1988), or the Arch model (Bass et al. 1991). This kind of modularity has been referred to as *interface separability* (Edmonds 1992).

At about the same interface separability approaches to user interface engineering were evolving, it was becoming more broadly apparent that different end-users needed different kinds of software support (Chin 1986; E. Rich 1983). The focus upon supporting users with different needs came to be known as *user modeling*. User modeling advocates the need for an embedded component to explicitly track and store information about its users (specific or prototypical), that would in turn be used support individualization of the interface. Research on *adaptive* and *intelligent user interfaces* first began to gather steam in this climate, and indeed some early work explicitly advocates combining the UIMS approach with user modeling (e.g. Hefley and Murray 1993).

Thus, adaptive and intelligent user interfaces were born out of a need for engineering solutions to the problems of interface complexity (see Hefley and Murray 1993) and individual differences between users (see Benyon 1993). In terms of the analysis presented in the introduction to this chapter, software needed to be flexible enough to accommodate many different users with different domain goals, but flexibility made systems harder to use and less targeted to a specific user's needs. Adaptive systems were developed to address the problem by manipulating the interface representation at runtime. The "first principles" of the engineering solution are:

1. Incorporate an abstraction that clarifies the separation between the interface representation and the rest of the system, and;
2. Add a component that can "observe" the user interaction domain at runtime, and modify the interface representation accordingly, so as to transform the user's interaction with the domain according to her needs.

In the early days of adaptive systems, much of the work with adaptive systems was focused around these principles. The definitions of adaptive systems offered by various researchers are illustrative of this. For Benyon (1993), the most important part of adaptivity was the separation between interface design and system functionality, and moreover, the specification of multiple designs to meet the needs of multiple users:

> "*Adaptive systems differ from other interactive systems in that they are characterised by design variety.... Rather than the designer trying to obtain a single solution to a problem, the designer specifies a number of solutions and matches those with the variety and the changeability of users and the environments. Adaptive systems are a serious solution to usability problems where a degree of variety is present."* (Benyon 1993)

Oppermann (1994), on the other hand, focused primarily upon the initiative the adaptive component might take in manipulating the interface:

> "*A system is called adaptive if it is able to change its own characteristics automatically according to the user's needs. The self–adapting system is the most common conception of adaptivity. Modification of interface presentation or system behaviour depends on the way the user interacts with the system. The system initiates and performs changes appropriate to the user, his tasks, and specific demands."* (Oppermann 1994)

Finally, Langley (1999), as a representative of the user modeling community, focused on techniques for gathering, representing, and processing information about the user.

> *"An adaptive user interface is a software artifact that improves its ability to interact with the user by constructing a user model based on partial experience with that user"* (Langley 1999).

The above aspects may be seen as dimensions of adaptivity that, at some point, have been critical areas of research within the adaptive systems community. For instance, Oppermann's observations form the basis for the rich body of work on mixed-initiative interaction (see Horvitz 1999), and Langley's view of adaptation is well supported by the user-modeling community. These research areas have their roots in a common underlying notion – that the *purpose* of adaptation is to transform the user's interaction with a system in a manner that brings it closer to the software interaction ideal.

Adaptive systems have since been developed for other sorts of reasons, but the bias towards the software interaction ideal is still present in much of the research that goes on. In the following, I provide a broad overview of the kinds of systems that have been explored.

2 Kinds of Adaptive Systems

Adaptive software has moved beyond simply helping the user interact more effectively with the system. Jameson (2003) offers a review which segregates adaptive software into that which is designed to improve the use of the system itself, and that which is designed to support "information acquisition." Information acquisition includes information retrieval systems, recommender systems, and tutorial systems.

In my own analysis, as in Jameson's, adaptive software that is concerned with the system itself accounts for a large and well-defined portion of the field. However, software that supports information acquisition is, in my analysis, part of a larger category of systems that support the users' domain activity directly, rather than supporting the user's use of the system. Often, the strategies employed by information retrieval technologies and recommender systems tend to be oriented to achieving the user's goal, based on some guidance from the user about how that goal is to be achieved, rather than attempting to manipulate the user's interactions.

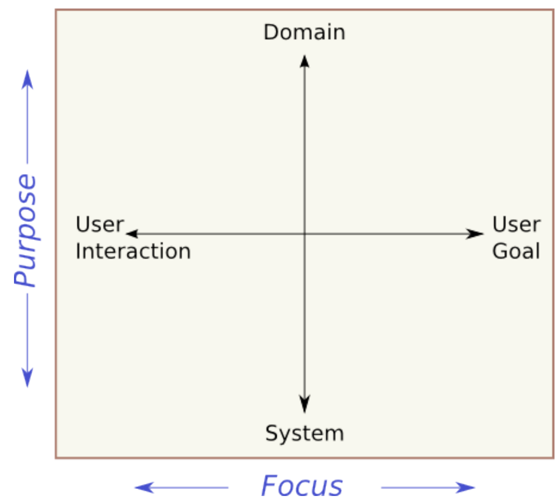


Figure 2: An adaptive software classification system

This analysis yields the two-dimensional classification shown in Figure 2. Along the horizontal axis, systems may be classified in terms of the which part of the user's activity the adaptive functionality is focused. On one end are adaptive systems which focus on manipulating the user's interactions, and on the other are systems that attempt to help attain the user's goal given some abstract definition of that goal. Along the vertical axis, systems may be classified according whether the purpose of the functionality is to improve the users domain activity directly, or if it designed to help the user use the system itself.

The location of any given system along either axis is a matter of degree. After all, a system is designed to support interaction in a domain, so facilitating the use of the system itself accomplishes a domain goal. Nonetheless, distinctions may be made between various systems, and I seek to do that in the following. It should also be noted that some systems include multiple adaptive strategies, and so occupy more than one quadrant in my system. I will note where this is the case in the examples covered.

3 System Oriented Adaptation

As discussed above, adaptive techniques were initially developed to cope with the tension between flexibility required to meet the needs of different users with a broad range of possible domain goals, and the software interaction ideal:

> **The Software Interaction Ideal**: *The user tells the software as directly as possible what they wish to accomplish in the domain, and the software just does it!*

The following two sections describes a variety of approaches that have been used to get a little closer to this ideal.

3.1 System Oriented Goal Adaptation

Adaptive systems that support the user's goal within the system itself typically work by advising the user about how to get to that goal. These systems approach the software interaction ideal by either eliminating gaps in the user's knowledge about how to accomplish something using the system (e.g. by using command X, and then command Y), or by telling the user about more efficient interaction strategies (e.g. instead of command X and Y, just use command Z). The addition of adaptive help makes sense when the underlying system is not designed with an API (application programer interface) that allows the software itself to be manipulated at runtime.

One of the first systems to do this, and one of the earlier adaptive platforms to be developed, was UC, the Berkeley Unix Consultant (Chin 1986). UC was an online help system designed to offer advice, answer questions, and help users debug problems at the UNIX command line. UC tracked the commands that the user issued, and from these deduced the user's level of expertise. This information was in turn used to tune the kinds of advice it might give the user.

A similar approach to facilitating interaction was taken with the Lumiere system (Horvitz et al. 1998). Lumiere was the research prototype that ultimately became the Microsoft Office '97 Office Assistant ("Clippy"). Like UC, Lumiere was designed to offer advice and help with user difficulties. Lumiere maintained a Bayesian user model, which was used to make inferences about a user's goals, their degree of frustration, and their expertise. Lumiere interacted with the user in several ways. If the user issued a query to the help system, this query would be analyzed and combined with the activity analysis to select help text for the user. In the absence of an explicit query, Lumiere would suggest courses of action if the likelihood the user needed assistance exceeded some threshold.

Both Lumiere and UC take the task domain to be the system itself, and both are concerned with improving use of the command language. Hence, the task model is bounded by the functionality of the system, and inferences need only be made about which command the user might want to use, and not about the parameters of that command. The only input the system requires are the commands the user issues, and possibly minor interaction with the user to dismiss help offerings.

3.2 System Oriented Interaction Adaptation

When it is possible for an adaptive system to interact with a piece of software more directly, manipulating the platform for the user can be a more straightforward way to reduce the amount of work the user has to do. This is perhaps the most well-explored type of adaptation, and is often taken to be an implicit goal in the development of adaptive functionality. Several approaches are presented in the following sections.

3.2.1 Eliminating Repetitive Action

One approach is to minimize the total number of interactions a user must take with the system. A way to do this is have the system try to notice certain kinds of patterns in the user's activity, and offer to complete the pattern. This type of adaptation was used in EAGER (Cypher 1991). EAGER was an agent-based programming by example (PBE) system for Hypercard on the Macintosh. EAGER was designed to reduce the mundane and repetitive activity the user had to do in the Hypercard environment. EAGER would monitor user actions for repetitive sequences, and when such a repetition was detected, it would offer to complete the user's task. For instance, if the user was copying a list of names from an address book into a document, EAGER might notice the repetition and infer that the user wanted to copy all of the names from the address book.

SmartEDIT (Lau et al. 2001) is another system that, like EAGER, is able to extract macros from sequences of repetitive activity with the system. SmartEdit is embedded in an Emacs editor. The user tells SmartEDIT when to start and stop recording actions to be generalized. Once this initial recording is done, the user can train the system further by stepping through the learned macro and guiding the system, or by providing further examples. When the user is confident the system has learned the macro, she can simply tell the system to execute the macro. These macros remain available to the user after definition. Extensions to the SMARTedit system have explored different modes of interaction with the user (Wolfman et al. 2001).

In systems like EAGER and SmartEDIT, the system is not actually inferring the user's goal, but is instead noticing patterns in the user's domain activity. This kind of adaptation is possible because the types of operations exhibited some kind of syntactic regularity the system could exploit. For instance, SmartEDIT's underlying inference engine made use of atomic operators that keyed off of the recognition of syntactic elements like words, sentences, and abstract ordinal positions with the text. The technique is powerful, and requires little interaction from the user, but it is only possible where repetitive patterns might emerge over sequences of syntactic operations in an interface.

3.2.2 Speeding up interaction

Another way to get little closer to the software interaction ideal is to speed up some of the user's interactions with the system, and so expedite the attainment of the user's domain goal. The general strategy is to reconfigure the interface so that the commands and / or information that will be used are easier to access, and the commands and / or information that won't be used are hidden. Note that this is precisely Benyon's (1993) definition of adaptive systems.

The Adaptive Toolbars technique is a very simple example of this (Miah, Karageorgou, and Knott 1997). The technique entails computing an “interestingness” index for operations that are represented by elements in the interface based on frequency of use. The system will then add or subtract operations them from a dynamic toolbar, and add or subtract other (non-dynamic) toolbars. A similar technique (SmartMenus) is currently integrated into the Microsoft Office™ family of products.

The Adaptive Toolbars approach requires very little in the way of input from the user – as long as the system can differentiate between users (e.g. via a login), the user's normal interaction with the system suffices. The adaptive functionality itself is “low-hanging fruit,” requiring little, but also offering little.

More powerful adaptations are possible in systems that are equipped with more information, such as with CHORIS (Computer-Human Object-oriented Reasoning Interface System; Tyler, Schlossberg, and Cook 1991). CHORIS was a middleware platform developed at the Lockheed Artificial Intelligence Center. The system used both task and user knowledge to infer an appropriate interface configuration for its user at runtime. CHORIS was deployed in several domains, including emergency crisis management and electronics manufacturing. In all cases, the system would automatically determine at runtime what subset of commands were required for a given user and task, and the system would dynamically reconfigure the interface as the user moved through his tasks. There were four dimensions of adaptivity within the interface: the location and detail of a map view, the kind of tasks available in a task list, the kinds of tools (like a chat window) that could be presented, and the way in which raw data was presented (e.g.

numbers or graphs).

The CHORIS system's adaptive functionality was driven by a combination of user modeling and intent inference techniques. CHORIS was equipped with an enormous amount of information about the roles its users could have, and the tasks that each of these roles entailed. The role-based user models were generated in advance of system deployment. Individual user models could be customized by the users to indicate a user's preference for the display of data, or which tools might be presented in the interface, but role assignments and task models were static.

Intent inference was possible in CHORIS because the interface was built around task models which were associated with each user's role. The task model itself was available in the interface, and every command issued by the user (by selecting from the task hierarchy directly, or by issuing a natural language utterance) could be mapped into the task model. This feature, combined with the fact that certain role definitions would not use certain tasks, simplified intent recognition to the point where it was a feasible adaptation strategy.

3.2.3 Automating an interaction sequence

Intent inference is not necessary for useful automation if the domain can be sufficiently constrained and the added adaptive functionality is not too disruptive. For instance, the LookOut prototype system was designed to add semi-autonomous scheduling functionality to the Microsoft Outlook email suite (Horvitz 1999). LookOut was designed to assist users with reviewing their calendar and scheduling appointments in response to incoming emails. When the user opened a message, the system would examine it for words or phrases that might indicate a meeting has been requested (e.g a date). Based on the precision and confidence of this assessment, the system would automate some of the interaction with the user's calendaring software. For instance, if the system could establish a proposed meeting time with high confidence, it would check the user's calendar and offer to schedule an appointment or propose an alternative. However, if the system could not establish a specific time for a meeting request, it might offer to bring up the calendar and select the most appropriate block of time for the user to review.

LookOut assumes that the user always has the goal of wanting to schedule a meeting. It will always identify times and dates, and will always try to interact with the user's scheduling software if it has sufficient confidence in its interpretation of the time and / or date. It may require some low-level input from the user. Whether or not it would become an annoyance would depend upon how frequently one actually wanted to schedule a meeting upon receiving some information about a date.

3.2.4 Elaborating Interactions

Sometimes, the best way to approach the software interaction ideal is to begin with that ideal, and introduce more work for the user only as necessary. This is a common approach in spoken language understanding systems, such as TOOT (Litman and Pan 2000). TOOT was a prototype spoken dialog system designed to retrieve train schedules. In began its interaction with a user using an ambitious discourse strategy that allowed the user substantial leeway in phrasing her requests. As misunderstandings occurred, TOOT would adopt a more conservative strategy, asking simpler questions so as to constrain the user's possible responses. Thus, rather than using adaptive functionality to approach an interaction ideal, TOOT's adaptive discourse strategy allows the system to be designed around a “best-case” interaction and for the interaction to degrade gracefully when the “best-case” cannot be achieved.

The preceding sections cover a wide range of adaptive systems that are designed to support interaction with the system itself. These systems have traditionally made up the bulk of the work in adaptive systems. However, other types of systems do exist, and have received more attention in recent years. These are covered in the following sections.

4 Domain Oriented Adaptation

In the following sections I review a few types of systems that are designed to extend a user's capabilities in a domain that exists outside of the software system itself. As above, systems might help the user attain some goal, or perhaps support the user's domain interactions more directly, but in both cases this support is oriented to the user's domain activity, rather than an interaction with the tool. In most of these cases, domain performance would be impaired or not possible without the adaptive functionality.

4.1 Domain Oriented Goal Adaptation

Systems that employ domain oriented goal adaptation help to achieve the user's domain goal given some abstract description of that goal, rather than manipulating the user's interaction to get them to their goal faster. I consider two types of goal-oriented adaptation here; that which is concerned with identifying information or resources for the user, and that which is concerned with creating some sort of product for the user.

4.1.1 Systems that seek information

Information retrieval systems make up one class of systems that use an abstract description of the user's interests to find some information they might want. For instance, the Adaptive Information Server (AIS) and Daily Learner were components of a web-based application designed to customize the delivery of news to users (Billsus and Pazzani 2000). The Daily Learner client interface provided the user with a list of current headlines, and several interface widgets to optionally indicate her degree of interest in a particular article as she read it. Based on this feedback, AIS built a model of the users' long- and short-term interests. This model was in turn used to customize news delivered to the user through a “personalized news” portal. By indicating her interests in news as it is delivered, the user provided the system with a description of her interests, and the system was then able to use this description to find other, similar, news items that might be of interest.

Recommender systems are another well-known class of systems that fit this schema. For instance, Recer (Chalmers 2002) was a browser-based application that monitored all file and URL access events within a given group of individuals, and used this information to recommend files or URLs to an individual based on their current browsing behavior. For instance, if an individual was browsing the website of a ski resort, Recer might display links to trail maps that others within the workgroup had previously browsed after looking at the same website. The underlying algorithm is a very simple collaborative filtering process, and the “description” that the user provides to the system is simply the document or URL they are currently looking at.

Note that Recer combines domain-oriented goal adaptation with a form of system-oriented interaction adaptation. It presumes that an individual's domain goal might always be to find related information about the currently viewed resource. Rather than make the user to go through the steps required to browse to this information (e.g. cutting and pasting a URL into the browser bar), it offers the user a shortcut in the interface for accessing it quickly. In this manner, Recer may be able to eliminate some steps for the user, whenever the user's domain goal is indeed finding related information.

Another system to offer a mixture of domain-oriented goal adaptation and system-oriented interaction adaptation was the AIDE platform (St. Amant and Paul R. Cohen 1998). AIDE was a mixed-initiative planning and navigation aid for exploratory data analysis (EDA). The interface was based on a direct manipulation metaphor, in which the user and system shared a common representation of the complete state of the problem via a graphical and tabular display of data. The user could select and manipulate this representation to perform various statistical operations. The adaptive component assisted the user during the natural course of the user's exploration of the data by inferring higher-level goals from the user's interface actions. For example, if a user requested a regression on a selected set of data, the system would run several different regressions to find the best fit, and identify outliers for the user. Control strategies encoded in the process provided sufficient opportunity for user-direction at each decision point in the exploration of the data.

The adaptive component in AIDE was driven by a script based mixed initiative planner which embodied standardized approaches to exploratory data analysis. The planner instantiated and executed plans based on user goals and variable bindings that were expressed via the user's interface activities. Goals and bindings were inferred where possible, but could be manually selected by the user as well. The user was able to interact with the underlying planner via a graph-based navigation panel, which portrays the task of data analysis as navigation through choice points in the process.

AIDE offered a form of domain-oriented goal adaptation in that it was designed to find statistical phenomena that would be of interest to the user. For instance, in one mode of operation, the user could tell AIDE to automatically move through its plan graph until it found something “interesting” to report to the user. At the same time, AIDE employed system-oriented interaction adaptation by reducing the number of interactions the user had to have in order to attain their domain goal. Thus, domain-oriented goal adaptation and system-oriented interaction adaptation go hand in hand within AIDE.

4.1.2 Systems that generate a product

Goal-oriented adaptive systems seek to attain some goal for the user, given description of this goal. In the case of the above systems, the “goal” is to find interesting information in the domain of interest. But in some cases, the algorithm contained within the system embodies a description of an appropriate product, and the user provides the system with information from which this product is to be composed.

For instance, (Perkowitz and Oren Etzioni 2000) describe a composite algorithm called “IndexFinder” to generate index pages that collect links to pages that users with specific needs might want to visit (e.g. links to similar products). IndexFinder works by analyzing usage patterns in a website, and conceptual categories that have been assigned to each page in a web site by the webmaster. Thus, IndexFinder functions as an adaptive web-page generation tool. The user in this case is the webmaster, and the user's inputs to the system are a set of conceptual tags over each of the pages in the site and a weblog containing traffic patterns. The system's output is a web page that meets the description of a “good index page” embodied within the algorithm.

Group decision support systems also fit this pattern. Group decision support systems combine multiple users' input according to some algorithm in order to make a choice that that is optimal with respect to some ideal. I will cover group decision support systems in more detail in Chapter 4, but an interesting example for purposes of this review is Masthoff's (2004) notional “personalization” system for group television watching. Masthoff acknowledges, but postpones, the consideration of several tricky problems, such as how the system would know who is watching, how individual preferences would be accumulated, and how other social considerations might intersect with the group decision to watch a given program on television. Her work does, however, go into depth in exploring how people in general combine their preferences, and studies how people respond to various approaches.

4.2 Domain Oriented Interaction Adaptation

Systems that exhibit domain oriented interaction adaptation are generally task specific, and hence do not generally infer the users' goals so much as they infer aspects of those goals. Several examples are provided in the following sections.

4.2.1 Automating Domain Activity

One way to support a user's domain activity is to automate some of the activities the user might have to perform in that domain. This idea is often invoked when discussing automated agents (Maes 1994). An example of this approach was the MASMA group scheduling platform (Cesta and D'Aloisi 1999). Like LookOut, MASMA was designed to make the task of scheduling meetings easier, however the system was designed as “stand-alone” software that was solely devoted to scheduling collaborative meetings. In MASMA, each user was associated with a meeting agent that maintained a model of the user's scheduling preferences and interests. Each agent was also assigned a degree of autonomy, which reflected the amount of trust the user had in the agent. All of these parameters were set manually by the user. Meeting agents would interact with one another to schedule meetings, requesting input from users at various decision points depending upon the degree of autonomy assigned to the agent.

MASMA does not need to perform goal inference, because it is a special purpose platform. It does, however, require an initial phase of interaction with the user to establish the user's preferences for a meeting, and depending on the user's comfort level with the automated scheduler, it may need subsequent interaction. User studies would be required to determine if the trade-off between the benefits offered by the system and the amount of work required to direct the agent would be worthwhile to the user.

4.2.2 Personalized Information Access

Techniques that have been used to simplify interfaces so as to speed up interaction in system-oriented adaptive systems can sometimes be transplanted into domain-oriented adaptive systems. As with their system-oriented counterparts, domain-oriented systems that employ such techniques attempt to reduce the number of interactions the user has with the system to get to the information they are interested in. Examples of this can be found in web personalization systems. For example, the AVANTI system was a web-based personalization system, designed with an eye to meeting the special needs of the elderly and disabled (Fink, Kobsa, and Nill 1988). AVANTI was implemented as a tourist information system, and changed types of information contained in web pages based on the needs and interests of users (e.g. handicap facilities at a hotel). All adaptations were based on a user model that contained information about the user's needs.

AVANTI's user model was populated by a questionnaire that was given to the user initially. A prototypical user model was chosen for the user according to the user's answers. The model was extended during the user's interaction to include likely user interests based on the user's browsing behavior. The user could also indirectly customize this user model by asking the system to show or hide different elements on each screen.

4.2.3 Reducing Information Overload

As domains mediated by software systems become more complex, one way a system can help the user out is to eliminate or simplify some of the information the user is confronted with while interacting wth a domain. This is often referred to as reducing "information overload" (Maes 1994).

Cummings and Mitchell (2007) offer an interesting example and analysis of this type of adaptation in an experimental system called MAUVE, which was designed to support the control of multiple UAVs (unmanned aerial vehicles). MAUVE was used to explore various levels of automation in a simulated UAV control domain. Several types of adaptive functionality were explored with MAUVE, corresponding to various degrees of automation.

In experiments with MAUVE, the user's task was to control four UAVs, each of which had several targets to eliminate. Targets were assigned to a user via an air-tasking order (ATO), which is the standard format for communicating targeting responsibilities. Each target was associated with a window of time – the time on target (TOT) – during which elimination of the target was possible.

The (simulated) UAVs were highly automated, and so were capable of executing pre-specified plans, but explicit control was required in several circumstances. UAVs could potentially require re-routing to avoid emergent threat areas. UAVs also required explicit human control during load, fire, and battle damage assessment phases of activity. Furthermore, because multiple UAVs might be scheduled to require user interaction at the same time, users might wish to postpone scheduled TOTs. To do this required an explicit request to the Air Operations Command. Requests were not guaranteed to be granted, but were more likely to be if made well in advance.

Three levels of automation were compared to a fully manual mode. In the manual mode UAV plans were arranged in a matrix format, corresponding to original ATO. The three levels of automation employed several different kinds of adaptivity. In the "passive" automation mode, the ATO was automatically interpreted and arranged in a timeline format; this was the only mode to deal exclusively with supporting the user's interpretation of information. The "active" mode added another level of support by highlighting “critical” areas in which multiple UAVs would require operator intervention at the same time, and also by automatically displaying a dialog so that the user could request postponing a TOT to deconflict plans. Finally, the “super-active” mode automated some of the user's activities by automatically issuing load and fire commands.

In MAUVE, some of the functionality included in the “adaptive” and “super-adaptive” modes was concerned with improving interaction with respect to system. Positioning a dialog in the interface is an example of “speeding up the interactions” (Section 3.2.2, above) and automatically approving UAV loading and firing requests are instances of “automating an interaction sequence” (Section 3.2.3, above).

However, automatically processing incoming information in some manner, and providing the user with an enhanced view of the domain is something else. The system transforms temporal data into a spatial dimension, so that the user might leverage her significant visual processing capabilities to recognize important patterns. Thus, the adaptive functionality manipulates the user's “read” interactions, but does this in order to improve the user's domain performance, rather than her interaction with the software. In this case, the system adds very little “intelligence,” but the approach is representative of a vector along which other adaptive systems might make substantial headway.

4.2.4 Intelligent Tutorial Systems

In general, intelligent tutorial systems represent another class of adaptive systems that use adaptive techniques, and manipulate the user's interactions with the system, but do so in order to achieve a goal that is outside the system. An example of this can be found in Andes, which is an intelligent tutorial system (ITS) designed to help students learn

Newtonian Physics (Conati, Gertner, and VanLehn 2002). It was intended to improve student's learning beyond that of students equipped with traditional study and homework materials. Andes uses Bayesian user models to make inferences about the student's knowledge and goals. Andes interacts with students through a problem solving interface by providing immediate feedback on partial answers to problems, and offering hints where errors are made or if the student requests help. Repeated hints for any particular problem are ordered, so that the first hint is the most general, and subsequent hints become more specific.

Andes elaborates interactions gradually, much in the same way TOOT elaborates its dialog with the user. Unlike TOOT, Andes manipulates the level of interaction that user has in order transfer knowledge effectively, in support of the domain goal of teaching the student.

4.2.5 Error Correction and Avoidance

Yet another common approach to adaptively mediated interaction in support of a domain goal is to provide support that either corrects or prevents errors. A very simple example of this is the automatic spelling correction functionality incorporated in most word processors, mail systems and web browsers.

Such adaptive technology can also be very useful when deployed in contexts outside the standalone computer. For instance, Adaptive Cruise Control (ACC; see Stanton, Young, and McCaulder 1997) is currently available in a variety of luxury automobiles. ACC uses radar or a laser based system to monitor traffic in front of the vehicle. If the vehicle is approaching another, it will slow down until it is possible to resume the driver's requested speed. ACC is sometimes coupled with a collision avoidance system that will alert the driver to stationary objects ahead, and apply the brakes and tighten seatbelts if no action is taken by the driver. This technology is an example of a safety-critical system, many types of which are currently deployed in many potentially hazardous situations.

5 Summary and Issues

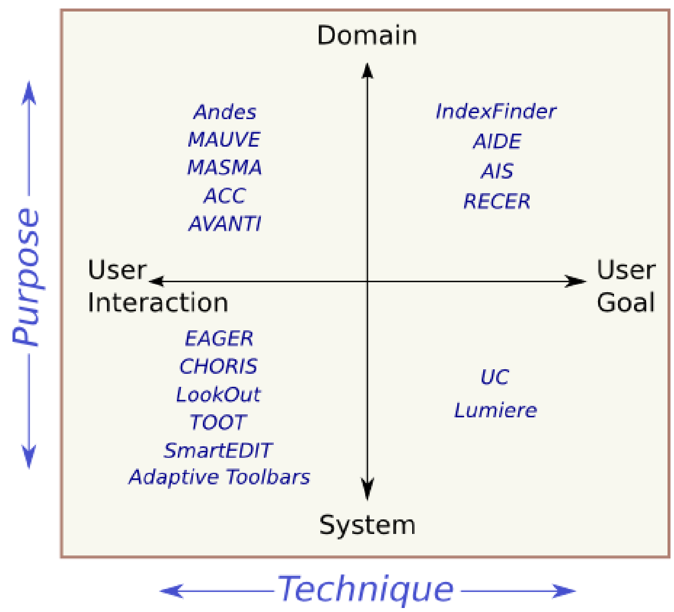


Figure 3: Location of systems surveyed within the proposed classification scheme.

The overview presented has covered seventeen systems, roughly arrayed within my proposed taxonomy as shown in Figure 3.

The systems described in the preceding sections were selected to illustrate the breadth of the field, and the techniques and hurdles they have in common. The survey does not reflect the number of different systems that have been developed within each category. For instance, although I have only mentioned two systems in the lower-right quadrant of Figure 3, adaptive help systems make up a broadly studied and well-developed approach to adaptive support.

In the following sections, I will make some observations that summarize two aspects of adaptive systems that have been introduced the preceding review. First, I focus on the issue of knowledge acquisition, which has been broadly acknowledged to be a hurdle wherever automation is employed. After that, I'll focus on the boundaries of the adaptive systems research, and discuss one avenue along which research might be extended.

5.1 The Knowledge Acquisition Problem

I have noted throughout this review where and how knowledge acquisition issues influence the power of a system, and how these issues have been addressed. Knowledge acquisition is a limitation on the power of an adaptive component. Jameson (2003) also notes the importance of user knowledge acquisition in adaptive systems, and Terveen (Terveen 1995; Terveen, Stolze, and Hill 1995) suggests that the degree to which the system's interface presents task relevant information will ultimately determine the scope of the adaptation. St. Amant and Young (2001) also note that it is easier to infer the user's domain intentions from their use of direct manipulation interfaces. This is because direct manipulation systems configure the interface to present domain tasks to the user as directly as possible. But as pointed out at the outset of this chapter, direct manipulation systems are not always possible, and when they are, they may not need much in the way of adaptive support.

One approach that has been used to get beyond such limitations has been to introduce “collaborative” agents into the interface, that communicate with the user about her intentions. One example of this is the COLLAGEN system, which is a middleware application designed to add intelligent, task-aware support to complex software platforms (Rich and Sidner 1998). COLLAGEN adds an interface agent to an existing software system. The agent is equipped with a plan library that describes how common domain tasks can be accomplished with the software. The agent communicates with the user, prompting her for the next step in a given domain task, and occasionally making suggestions or performing domain activities if the target platform allows this. Explicit communication between the agent and the user occurs via a structured language, which allows the user to direct the system through the plan library at given choice points. Subsequent versions of COLLAGEN added plan recognition capabilities, so that the agent could infer the user's goals without the user's explicit direction (Lesh, Rich, and Sidner 1999).

COLLAGEN is an interesting example for several reasons. As with CHORIS, and AIDE, it is equipped with a full domain plan library, and seeks to support users' activities using this library. However, because it is designed to bring help to an existing application, the software interface may or may not be designed around task operators in the domain. Thus, accurate plan inference in the general case is likely to be hard, especially when trying to support novices, whose activities at the operator level may or may not correspond well to actual operational sequences in the task domain. COLLAGEN solves this problem by engaging the users in a conversation about the domain, using an artificial discourse language.

Several observations have been made about COLLAGEN that suggest that this is not a viable strategy. As was noted by COLLAGEN’s developers, “it is often more efficient and natural to convey intentions by performing actions" (Lesh, Rich, and Sidner 1999; page 23), and this was the reason for including plan recognition in subsequent systems. But even with plan recognition, the dialog strategy employed by COLLAGEN may be too restrictive to support the kind of fluidity that is natural for users. As pointed out by (Ferguson and James F. Allen 1998), any revision to a plan within COLLAGEN requires backtracking through the hierarchy in order for the system to maintain context. This is likely to be a tedious engagement for the user.

Generally, attempts to move beyond these limitations by adding conversational agents to the interface have not met with much success, for a couple of related reasons. First, a conversational agent will never be an "equal" partner in the conversation until natural language interpretation is perfected (see Alterman 2000 for more on this). As a result, any conversational engagement between the user and agent will require extra work on the part of the user to make herself understood. Consequently, any interaction with a conversational agent distracts from the user's domain goal, because it requires the user to focus her attention on interaction with the system. It is a "breakdown" that detracts from the tool aspect of the system itself (Bødker 1990).

Research in spoken language, multi-modal, and affective interfaces are all designed to find new ways to get information into the system without distracting users from their domain task. But new techniques are always of value, and I provide one that can be applied in groupware systems in the final section of this chapter.

5.2 The Boundaries of Adaptive Systems Technology

The majority of adaptive systems are system-oriented. This is not surprising, because the communities that study adaptive systems – such as the intelligent user interfaces community and user modeling communities – have their roots in this type of adaptation. There have, however, been criticisms of this approach as a technique for system design (see Maes and Schneiderman 1997). In these criticisms, adaptation is considered to be an ineffective, or overly complex solution to a problem that might be better addressed with more effective design.

Unfortunately, while this blanket criticism may be warranted in some cases, it obscures some of the novel capabilities that adaptive techniques have to offer. There is a difference between using adaptive technology to solve a design problem, and adaptive technology that is designed to augment human capabilities. Domain-oriented adaptive systems are systems that do the latter.

The largest blanket differentiator between domain-oriented and system-oriented adaptive systems is what these systems are adapting *to*. With system-oriented adaptive systems, the purpose of the adaptive technology is to achieve the software interaction ideal. In domain-oriented adaptive systems, the purpose of the adaptive technology is designed to achieve an ideal that is relevant to the domain. Error-correcting systems might been seen as the embodiment of the ideal of "perfect action," information retrieval systems that of "perfect knowledge."

There has not been a focused effort to identify and elaborate these ideals. Much as with the software interaction ideal, they are implicit in the design of systems, and very often taken for granted. But the ideals that underlie existing systems can be enumerated, and limit the variety of adaptive systems that are typically considered. By making a conscious effort to develop and cultivate new ideals of human activity, we can broaden the set of avenues along which adaptive systems might be developed.

In the following section I discuss approaches to the two issues with adaptive systems that have been raised in this section and the preceding one. First, I will describe a technique for knowledge acquisition that becomes possible in groupware systems. Following that, I will discuss a type of adaptive system that has not yet been broadly explored, but is representative of an ideal that presents a rich avenue for future investigation. Both of these aspects depend upon the fact that communication in groupware is mediated by the system itself. These two topics are the focus of the remaining chapters in this dissertation.

6 Mediated Communication and Adaptive Groupware

Groupware is software that is designed to support groups of people engaged in a task (Ellis, Gibbs, and Rein 1991). As with other software, groupware mediates the users' interaction with a domain in pursuit of a goal. However, a unique aspect of groupware is that some of the interactions the users have are not with a task domain, but with other users. So, while groupware mediates interaction with a task domain (like single-user software), it also mediates the interactions with other users.

Mediated communication can have both good and bad impacts on people's ability to coordinate their activities (including their conversation). Mediated communication lacks the immediacy and context that people depend upon in face-to-face communication, Clark and Brennan (1991) describe how various dimensions of mediated communication can impact the maintenance of common ground, both in positive and negative ways.

However, it presents the adaptive system developer with a couple of opportunities.

The information that people pass through the system will, to some degree, be information that is necessary for them to stay coordinated in their task. Thus, information that might be leveraged to drive intelligent support is already available in the communication that goes on between coordinating users. Unfortunately, most user communication occurs via unstructured channels such as chat or message boards, making it difficult for the system to interpret.

A potential solution to this problem is to introduce a mediating structure into the communication channel that will make the necessary information available to the system. It is critically important that this mediating structure be useful to the user, and not simply be included in order to support knowledge acquisition. The following chapter will be devoted to demonstrating that such mediating structure can be identified, and that it can indeed provide valuable information for an adaptive component.

In addition to creating an opportunity for knowledge acquisition, the fact that the system mediates communication between users also means that the collaborative process itself can become a target for adaptation. This kind of adaptation is representative of a different sort of ideal that has not been broadly explored within the adaptive systems community. I'll refer to this form of adaptation as *adaptive mediation*.

One of the best examples of this approach can be found in a proposed system called PIERCE, which was designed to function with the Epsilon learning environment (Soller 2004; Soller and Lesgold 1999; Goodman et al. 2005). Epsilon combined an intelligent tutorial system with a collaborative learning environment, to provide a virtual space where students could interact over course materials. Techniques were developed within Epsilon to monitor the students' interaction, and detect and classify potential dysfunction in the collaborative process. Goodman et al. (2005) describe a prototype agent called PIERCE, which was designed to use this knowledge to "jump in" to the conversation whenever negative episodes were detected and attempt to fix things with a targeted contribution to the collaborative dialog. No follow-up studies have been published at the time of this writing.

The algorithms that drive the proposed Epsilon / PIERCE system embody an ideal collaborative process, based upon literature and analysis of data collected from collaborative learning episodes. This ideal is used to guide the collaborative process to something more like it, in support of the domain goal of more effective collaborative learning.

A drawback with Epsilon is that in order for the system to acquire the information necessary to drive the adaptive functionality, all posts made by users to the collaborative forum must be tagged as one of several different kinds of speech act (Searle 1969). It is not clear how onerous this is for the users, but it might be addressed through the incorporation of useful mediating structure.

The remainder of this dissertation is devoted to elaborating upon the two opportunities that have been identified. In the following chapter (Chapter 3), I will describe a groupware system that employs a mediating structure that reduces user work, and provides the basis for adaptive functionality that is both system and domain oriented. Chapter 4 will provide a review of literature on decision making and group decision making which will be necessary to contextualize Chapters 5 through 7. Those chapters will describe a system that uses the knowledge acquisition technique described, but also employs adaptive mediation to guide collaborators to a notional ideal, and away from a well known problem in collaborative information processing. Following this study, I will offer some final observations, and a summary of the territory that has been covered.

Chapter 2: VesselWorld Case Study

As discussed in the preceding chapter, acquiring enough run-time information about the user to provide intelligent support is a central issue in building any adaptive system (Fischer 2001; Jameson 2003; St. Amant and R. Young 2001). To tutor a student, a system needs to know the student's strengths and weaknesses; to offer context sensitive help, it needs to know what the user wants to accomplish; to help find relevant content, it needs to know the user's interests. For a collaborative system to adapt to its users' needs, it must have information that is relevant to collaboration – information like the users' shared goals, common knowledge, or roles.

A major hurdle in the design of a user knowledge acquisition strategy is how to encourage the user to furnish the information that the system needs without impairing the overall usability of the system. There are numerous approaches to the problem. Mixed initiative systems (e.g. Maes 1994; Horvitz 1999) often obtain information through agents that engage the user to jointly solve a domain task (e.g. TRAINS, Ferguson, Allen, and Miller 1996; LUMIERE, Horvitz et al. 1998; COLLAGEN, Rich and Sidner 1998). Intelligent tutorial systems exploit the information gained during the testing process to derive models of students (e.g. Anderson et al. 1995). Multi-modal systems (e.g. Bosma and André 2004) and affective interfaces (e.g. Kaiser 2005) integrate other sources of information, like gestures and biometrics, to enhance the information available to a back-end inference process.

In groupware systems, much of the information required to drive intelligent support is already available in the communication that goes on between coordinating users. Unfortunately, most user communication occurs via unstructured channels such as chat or message boards, making it difficult for the system to extract useful information. One way to address this is to introduce structured communication channels (e.g. Malone 2001; Soller 2004). When introducing structure, however, care must be taken so as not to impair the users' ability to communicate with one another.

In this chapter, I describe a method for introducing structure that people will use, that helps them stay coordinated, and at the same time provides useful information to the adaptive system. The method hinges upon the development of task specific communication tools that support the exchange of structured information relevant to coordination. Because these artifacts add structure to communication, the information that passes through them can be easily exploited by runtime algorithms. Because the information is relevant to coordination in the task domain, it will be useful for domain-specific intent inference.

The work reported on here is the result of a collaborative effort that has examined the development and use of mediating artifacts in groupware environments (Alterman 2000; Alterman et al. 2001; Feinman and Alterman 2003; Landsman and Alterman 2005; Introne and Alterman 2006). Here, I present a series of case studies demonstrating the described knowledge acquisition approach in a testbed groupware environment that served as a platform for much of the collaborative research effort.

1 Coordinating Representations

There is a rich body of ethnographic research that describes how people design and use structured artifacts to support specific and complex coordination as part of an established work practice (Suchman and Trigg 1992; C. Goodwin and M. H. Goodwin 1996; Hutchins 1995a; Schmidt and Simone 1996). These kinds of artifacts have several common properties. They are external representations that address recurring problems of coordination within a community. They make it easier for individuals to align their private views of the world. There is a protocol that describes their typical use within a particular community of practice. Following Suchman and Trigg (1992), we refer to these kinds of artifacts as *coordinating representations* (CRs) (Alterman et al. 2001).

There are many examples of CRs in everyday life. The stop sign is a CR that helps people to coordinate their behaviors at an intersection. The arrivals and departures board at a train station supports coordination between many people and the operators of a railway terminal. The grocery list on the wall in the kitchen helps a couple coordinate the activity of supplying a household. The stock market "ticker" helps millions of people coordinate their market activities.

There are also many examples of CRs that have evolved over time to support specialized communities. For instance, the *complex sheet* in an airport setting is an example of a paper-based coordinating representation that helps coordinate transfers of baggage and people between gates in an airport during "complexes" (Suchman and Trigg 1992). Complexes are pre-scheduled periods of time during which all gates fill with incoming planes, transfers are made, and then all planes depart. As shown in Figure 4, the complex sheet is a matrix mapping incoming to outgoing planes, and cells in the matrix are used to indicate the transfer of people or baggage. Planes are ordered chronologically along both axes, so completed transfers are checked off diagonally downward and to the right across the cells of the matrix.

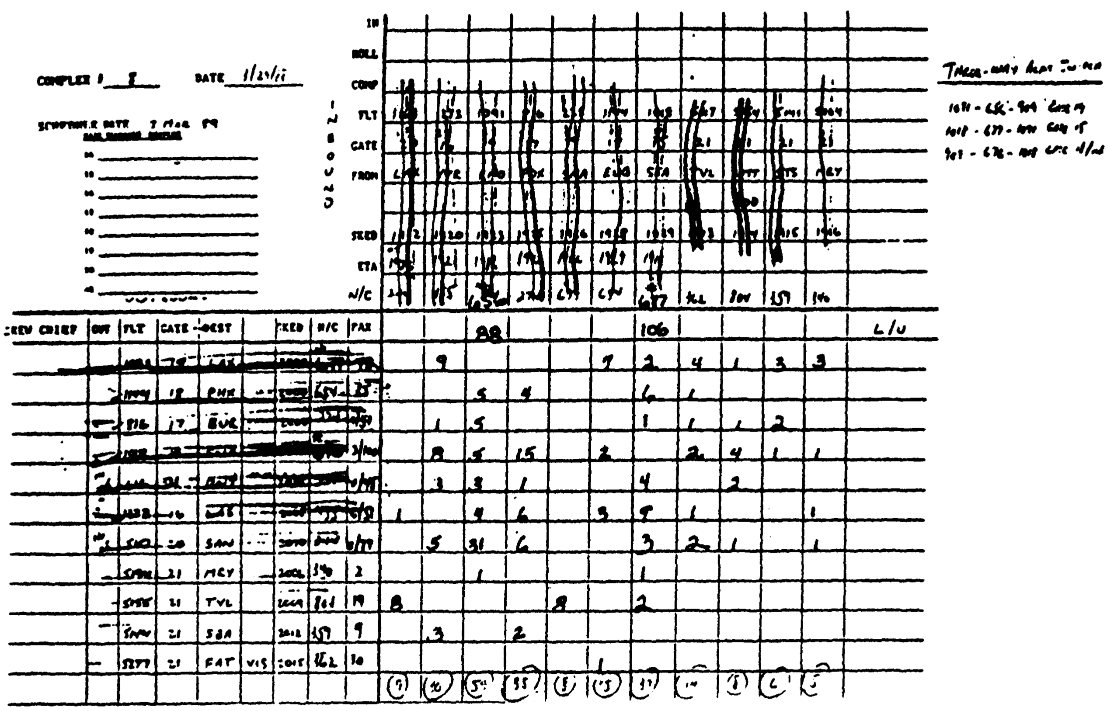


Figure 4: The complex sheet (Suchman & Trigg, 1997)

As computers have become common in work settings, computer-based coordinating representations have been developed. The bug-report form is a CR that is used by software engineers to structure and monitor the progress of a software engineering process. Shared calendars, call tracking databases, and inter-office wikis are all examples of general CRs that can be used to share information and improve coordination within a multi-user environment.

When CRs are introduced directly into a system's communication channels, we are able to enhance coordination both via the intrinsic usefulness of the CRs and also by gathering information that a groupware system can then use to provide intelligent support. This chapter discusses evidence of such expanded utility in one system and describes a methodology that can be applied to develop adaptive components in other groupware systems.

1.1 Organization of the argument

This work primarily concerned with how to acquire knowledge about users and their at runtime context to drive intelligent algorithms, without creating unnecessary or undesirable work for the users. The solution proposed is to modify the representational properties of the software system so as to both improve user coordination and gain access to the necessary runtime information. Hence, a single design solution is used to improve the system from the users' perspective, and from the perspective of the adaptive component designer.

To serve as a roadmap, an outline of the argument is provided here as a series of questions and answers.

1. **Question**: How can adaptive support be added to a groupware system?

   Answer: We infer users' intentions.

2. **Question**: Where can the run-time user information required for intent inference

be obtained?

Answer: This information is available in unstructured communication channels, but because it is unstructured it is difficult to access.

3. **Question**: How can this information be transformed so as to make it available to the system?

   Answer: By introducing a structured communication channel in the form of a coordinating representation.

4. **Question**: If users are provided with this more highly structured communication channel, will they use it?

   Answer: Yes, because the structure that is provided by the coordinating representation makes it easier for users to exchange some kinds of information.

5. **Question**: Can a coordinating representation be developed that does not impair user performance?

   Answer: Yes; in fact, CRs improve domain performance.

6. **Question**: Is the information that becomes available via the use of the CR useful for intent inference?

   Answer: Yes, because it offers a window onto task and coordination-specific knowledge that is shared by users, and this knowledge is highly structured.

The results from case studies will be presented that confirm the answers to the above questions, demonstrating that:

1. Users choose to use the coordinating representation over chat to exchange some kinds of information (*Section 4.1*)

2. Using the coordinating representation helps users perform their task better. (*Section 4.1*)

3. The information provided by the coordinating representation enables intent inference with good accuracy. (*Section 4.2*)

4. An adaptive component that is driven by the information made available is heavily used, improves performance in the domain task, and reduces cognitive effort. (*Section 5*)

The chapter proceeds as follows. First, the domain, groupware platform, and associated intent inference technique are presented. Some of the representational deficiencies of the groupware system are then described; first from the perspective of the user, and then from the perspective of the developer of an adaptive component. It is then shown how a CR addresses these problems. Empirical data is presented that demonstrates how the CR benefits the users, and how information derived from its use improves intent inference. An adaptive component that uses information from the CR is described, and further empirical data confirming the utility of the adaptive component is presented. The chapter will conclude with a summary of the overall methodology, some final thoughts about the generality of the approach.

2 VesselWorld

The experimental platform used in these studies is a groupware system called

VesselWorld, in which three participants collaborate to remove toxic waste from a harbor. VesselWorld was demonstrated at CSCW 2000 (Landsman et al. 2000). The platform was designed with the goal of studying the coordination problems commonly faced by groupware users. The domain task requires varying degrees and types of coordination, collaborators have different roles and responsibilities, and coordinating information is exchanged via a chat window. These features are highly relevant to the study of groupware systems in general. VesselWorld proved to be very challenging for its users; in our case studies, we found that the performance of a given user group usually did not stabilize until after roughly seven hours of use (including a two-hour training session).

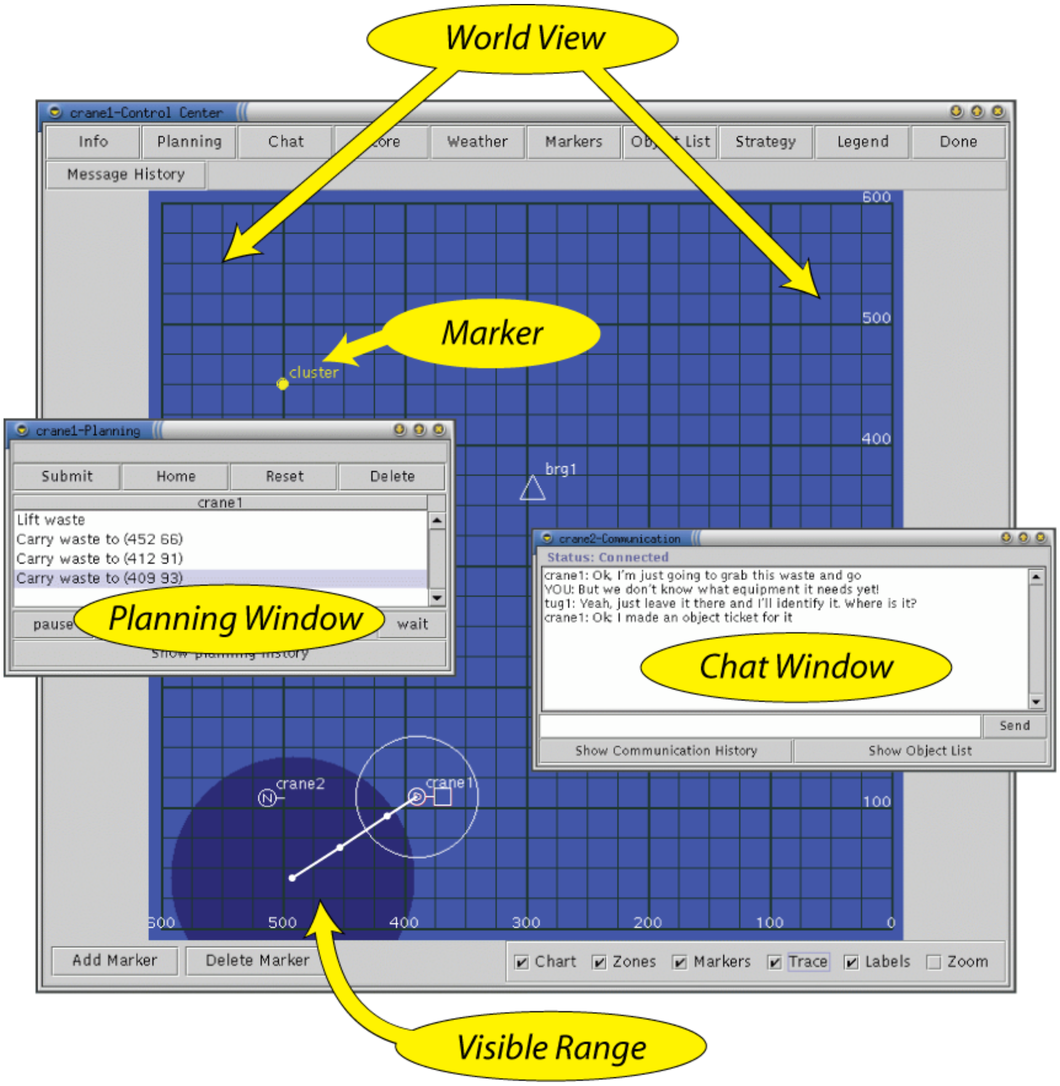


Figure 5: The VesselWorld system

VesselWorld presents a relaxed WYSIWIS environment, shown in Figure 5, in which three participants play the role of ship's captains, and their joint goal is to remove toxic waste barrels from a harbor without spillage. The main interface ("World View" in the figure) is a shared map; the x (increasing to the west) and y-axes (increasing to the north) indicate latitude and longitude respectively. Users are provided a private "marker" facility so that they may annotate the world view with pertinent information (see the "Marker" in the figure). Each ship can only “see” a small radius around its current location (the darker circle marked “Visible Range” in the figure), so each user has different directly observable domain information; hence users must communicate to maintain shared information about wastes in the environment. Communication may occur at any point, but all communication occurs through a text-based chat window that is part of the system (“Chat Window” in the figure). VesselWorld logs complete interaction data that can be used to replay user activity. This later feature is discussed further in Section 6.

The progression of a VesselWorld session is turn-based, such that every user must submit a single step to the server before it evaluates them and updates the world on each client screen. Users may plan any number of steps in advance, although each step can only involve objects that are currently visible. Plans can be managed (steps may be deleted or plans reset) via a separate planning window (“Planning Window” in the figure). Users’ plans are not visible to each other, again requiring explicit communication to manage coordinated plans.

A VesselWorld session is complete when all toxic waste barrels have been moved to a large barge, which has a fixed position and unlimited capacity. Each ship has a different role in this process. Two of the ships have cranes that can be used to lift toxic waste barrels from the harbor and load them onto a barge. The third user is a tugboat that can be used to drag small barges (which have limited capacity) from one place to another. The crane operators can load multiple wastes onto the small barge, and at least one of them must also be present to unload the barrels and place them on the large barge. For notational convenience, we will occasionally refer to the crane operators as “cranes,” the tugboat operator as the “tug,” and toxic waste barrels as toxic wastes or simply wastes.

Wastes are of different types and require different coordination strategies to be removed from the harbor. A single crane may lift a small or medium waste, but two cranes must join together to lift and carry a large waste, and an extra large waste may be jointly lifted but can only be carried on a small barge by the tug. Wastes may require specialized equipment to be moved, and the cranes carry different types of equipment. The tug is the only actor who can determine the type of equipment a waste requires.

The users are scored by a function that takes into account the number of barrels cleared, the number of steps this took, the number of errors (dropped waste barrels) made, and the difficulty of the problem. In all user studies, the users were instructed to try to maximize their score.

2.1 Intent Inference in VesselWorld

Planning in VesselWorld is a laborious and error prone operation (Alterman et al. 2001). User errors often occur because of forgotten plan steps or joint plans that have become unsynchronized. We sought to develop an automatic plan generation tool to address these problems. A hurdle in making such a tool useful is that there are an overwhelming number of potential goals for each user at any given time. Thus, an intent inference procedure was developed to reduce the number of possible goals to a manageable list from which users could then make a selection.

Bayesian Networks (BNs; Pearl 1988) were used to infer user intensions. BNs have been used in numerous systems and in a variety of ways to perform plan recognition (e.g.

Charniak and Goldman 1993; Horvitz et al. 1998; Albrecht, Zukerman, and Nicholson 1998). The approach taken here is straightforward. The BNs used in VesselWorld are static and model the relationship between information about the state of the domain and users' likely intentions. At runtime, information about the domain is posted to these models for each agent-waste pair separately. The likelihood an agent has an intention with respect to a given waste is read off one of the nodes in the network, and the intention with the highest value for a given agent-waste pair is taken to be the most likely one for that agent. New information is posted to the BN whenever a relevant change in the world is detected.

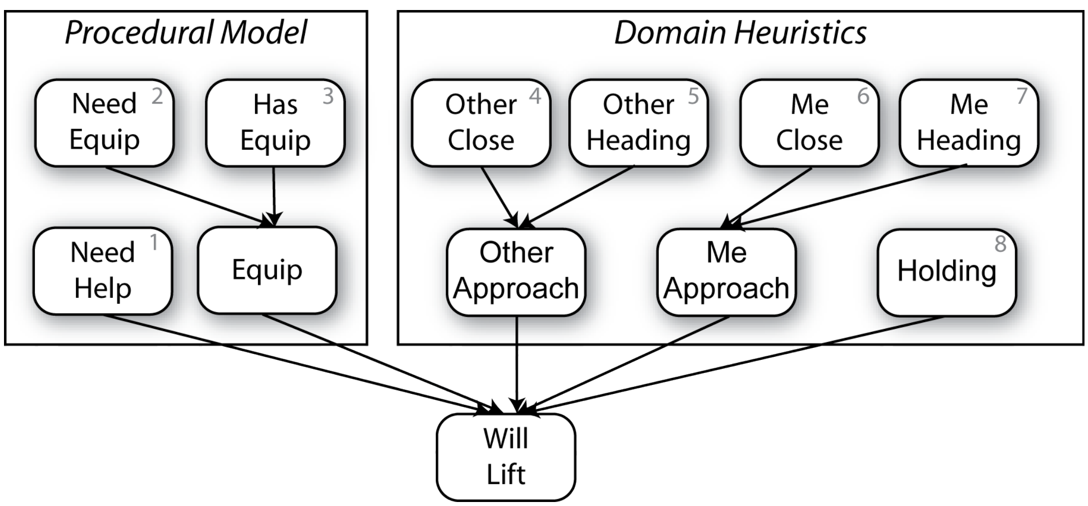


Figure 6: Schematic of BN used to infer Crane Lift Intentions (information is posted to numbered nodes)

Two BNs were developed, one which infers crane intentions and one which infers tug intentions. Together, the two BNs can infer seven goal types; JOINT LIFT (<crane><waste>), LIFT (<waste>), JOINT LOAD (<crane><waste>), and LOAD (<waste>) for the cranes, and BRING (<small barge> <waste>), ID (<waste>), and TRANSFER (<small barge ><large barge>) for the tug. In this paper, we restrict our analysis to the portion of the crane network that predicts the Cranes' intentions to lift wastes. This BN is shown in Figure 6; it models the likelihood that a crane operator has the intention to lift (or jointly lift with the other crane operator) a specific toxic waste.

The information used to determine if a crane has an intention to lift a waste is:

- The size of the waste (which determines whether a single crane can lift the waste, or the support of another crane is required). (*Node 1*)
- The type of equipment required by the waste and the type of equipment the crane has. (*Nodes 2 & 3*).
- Whether the cranes are close to or heading towards the waste. (*Nodes 4-7*).
- If the crane actor is currently holding a waste. (*Node 8*).

As portrayed in Figure 6, the BN combines a procedural model of the task-domain with domain-specific heuristics that indicate which of those actions are most likely. For example, a procedural constraint is that an individual crane cannot lift a waste unless it either has the right type of equipment, or the waste does not require equipment; this information is captured in three nodes, "Need Equip," "Has Equip," and "Equip." These three nodes do not represent a simple rule, because the equipment requirements for a waste might not be known, and hence it is necessary to explicitly handle this kind of uncertainty.

Heuristic factors interact with the procedural model to differentially weight those wastes that are possible lift candidates. For the cranes, these factors are how close the ship is to a waste, how close the current heading of the ship is to a path that will intersect with a waste, and whether or not the operator is already holding an object. This heuristic information in part reflects the physical constraints of the domain; a ship must be close to a waste before it can be lifted.

As a whole, the BN models users that have a certain degree of proficiency with the domain. It is expected that such users understand the rules of the game, and will only try to lift wastes that they can. It is also expected that these users will proceed in a generally predictable fashion. That is, they will not zig-zag wildly upon an approach to a domain object, and they will not randomly pick up wastes and put them down again. It would be possible to incorporate additional models for less proficient users, or model users' domain concepts more precisely (e.g. Horvitz et al. 1998). The focus of this work, however, is upon solving the difficulties in acquiring the user information necessary to drive the user model at runtime.

3 Representational problems in VesselWorld

Any piece of software may be considered to be part of a representational system. A representational system has three essential ingredients (D. A. Norman 1991):

1. The represented world (that which is to be represented);
2. The representing world (a set of symbols);
3. An interpreter (which includes procedures for operating upon the representation).

The representing world mediates the interaction between the interpreter and the represented world, and in so doing constrains the types of interaction that are possible between the two. As applied to software, the software system is a representational medium that constrains both how information in the represented world is viewed, and what procedures are available to the user for interacting with this world.

For adaptive systems, software can be seen to function as a mediating layer in two distinct but intersecting representational systems. First, and most clearly, it mediates the interaction between the user and domain (which may include other users), transforming the represented world so as to (hopefully) better support domain activity. The software system also mediates the interaction between the designer of the adaptive component and the user. In this later case, software very much constrains what information the designer can acquire about the user at runtime.

Here, representational deficiencies in VesselWorld are considered from both perspectives. The problem from the user's perspective, which is most traditionally thought of as an HCI problem, is considered first. The problem from the perspective of the adaptive component designer, which is a user knowledge acquisition problem, is then described. Both these kinds of problems are exactly those which Coordinating Representations can solve. Following the analysis of these problems, it is shown how the introduction of a CR in VesselWorld does just that.

3.1 Problems from the user's perspective

During a VesselWorld problem solving session, users must search for and discover information about the waste barrels that are floating in the harbor. Because each user has different locally available information, and recovering wastes requires the coordinated efforts of multiple users, it is necessary that participants communicate to establish mutual knowledge about the wastes. Managing and maintaining this shared information comprises a significant portion of the work that users must do during a VesselWorld session.

VesselWorld initially provided two tools to support this work. The chat window allows users to communicate information about objects in the domain. It provides a scrollable history that records the entire dialogue during a given chat session, so that users can more easily recover information that may have been missed or is forgotten. The other tool is a “marker” facility, which allows individual users to annotate their own private maps (each user can only see their own markers). A marker consists of a highlighted point on the map, and a free-form text message for recording information (see Figure 5).

These tools reflect the designers’ initial vision of users’ representational needs for managing domain information and establishing mutual knowledge about waste barrels. It was expected that users would publish local information via the chat window, and use the marker facility to record all waste information. During actual system use, it was found that these tools did not provide users with sufficient support. These representational deficiencies were most clearly manifest in specific features of the users’ runtime dialogue.

One such feature was explicit talk about coordination, as shown in Figure 7. In the example Crane2 suggests that the participants work out a shorthand for referring to waste information (line 3). Such explicit talk reveals a perceived need by the users for some form of coordinating structure.

```
1. Tug1: There are two waste sites near me,
   one at 120/415, one just SE of it.
2. Crane1: what sixze/ equi[p?
3. Crane2: hmm shall we come up with a short
   hand for a waste info?
4. Tug1: The Se one is small, needs a
   Dredge, the first is small, and need NO
   equipment. Does that mean I can pick it
   up?
```

Figure 7: Explicit talk about coordination

Another feature of user chat was the emergence of explicit conversational patterns for managing domain information. For example, one group would perform a "marker check" when there were discrepancies among individual's local waste information, in which all of the known wastes would be validated. First, an actor would announce a marker check, and then proceed to transcribe all of her local waste information into the chat window. The other actors would confirm or contradict this information according to their own private views of the world. The transcription process that this group of actors undertook involved a significant amount of mundane work. However, users made the decision that this was less work than trying to recover from errors that were caused by inconsistent information about the world.

Recurring coordination problems were also indicative of a representational deficiency in the system. Frequently, one actor's reported perceptions raised questions from another actor, possibly about the size or location of a waste. This was usually caused by a lack of precision in communicating about locations, leading to difficulties in resolving references to objects. Sometimes, these errors would lead to one actor moving to the area of the waste in dispute to resolve the problem.

The coordination problems and dialogue features described above can be generally described as grounding problems (H. H. Clark 1996). The creation and maintenance of common ground forms a substantial portion of the work that goes into conversation (Ferguson, James F. Allen, and Miller 1996; Traum 1995). Research has also demonstrated that technological media have impacts on grounding (Clark and Brennan 1991; Brennan 1998; see also Clark and Wilkes-Gibbs 1990).

The analysis methodology described here builds upon some of this earlier work on the grounding process (see Alterman et al. 2001 and Feinman and Alterman 2003 for more detail). This analysis elucidates those coordination problems that lead to the design of domain specific CRs, and grounding problems are one type of coordination problem that CRs can address. Critically, CRs solve two problems in the design of adaptive groupware. They address coordination problems for collaborating users, and they can also address a knowledge acquisition problem. In the following, the knowledge acquisition problems in VesselWorld are examined more carefully.

3.2 Problems from the adaptive component designer's perspective.

The BN introduced in Section 2 is a model of the behavior the designer expects of users in different situations at runtime. For instance, the designer expects that if the two cranes are approaching a large waste and are not carrying anything, there is a good chance they will attempt to jointly lift the waste. In order for this model to function, information about the user and the user's runtime context is required. Some of this information can be derived from the plan steps users submit (where the users are, what direction they are heading in, what they are holding). However, a significant portion of the required information about wastes (where they are, what size they are, what equipment they require) is only available in the users' discussion about the domain. This information is very hard to extract automatically.

Figure 8 is a portion of the chat logs taken from a typical planning session in VesselWorld. Note that users sometimes refer to toxic wastes by coordinates on the shared map (e.g. "105, 420"). In the first line of the example, Crane2 announces a waste at (120, 420). In lines 2-4, Crane1 asks for clarification about the specifics of the waste. In lines 5-6, the tug replies (having apparently already investigated that toxic waste barrel) with corrected coordinates (105, 420) and specific information about the waste. In line 8, Crane2 thanks the Tug operator for the clarification, and the Tug closes the conversational turn in line 9.

```
1. Crane2: I found a waste at 120
   420
2. Crane1: ok
3. Crane1: what type of waste?
4. Crane1: large,small?
5. Tug1:    105 420 needs a dredge,
   i think that is where you are
6. Tug1:   small
7. Crane1: ok
8. Crane2: Thanks for checking
9. Tug1:   no problem
```

Figure 8: Excerpt from chat during VesselWorld session

This dialogue illustrates some of the problems in automatically extracting the domain information required as input to our intent inference procedure. In order to associate information appearing in separate utterances with a single concrete waste, it is necessary to correctly resolve references. However, because the dialogue occurs between three active participants, the conversational turns that might be used to narrow the reference resolution scope are hard to identify. Furthermore, referring expressions can change from utterance to utterance even within the same conversational turn. For example, line 1 refers to the waste as “120 420” and line 5 refers to the same waste as “105 420.” People can sometimes handle such ambiguities, but this is problematic for automatic reference resolution algorithms.

Within user groups, some structural conventions for referring to domain objects do emerge, but the type and use of this structure varies widely between and within groups. Some groups are explicit in defining a shorthand; other groups converge to common conventions over time. Within groups, referential styles vary between participants and over time. Oftentimes, multiple conventions will co-exist. In Figure 9, a section of a transcript is shown for a group that exhibits this behavior. In the segment of dialogue shown, the players are announcing waste barrels they have found. In this group Crane2 always includes an “@” between the type and the location portions of the referring expression, and a question mark is used to indicate that equipment requirements for the waste are unknown. Crane1 usually places the type description before the location, and the tug places a type description after the location. The participants never converge to a common convention

```
1. Crane1:      l 543 204
2. Crane1: s 562 150
3. Crane2:      X?@150,559
4. Tug1:          1 395 lg dredge
5. Crane2:      s?@190,434
6. Crane2: s?@202,336
7. Crane1:      sm 394 71
8. Crane1:      large 395 22
```

Figure 9: Different ways of referring to wastes, within one group

For purposes of knowledge acquisition, the problem with such inconsistencies is that it is very difficult to produce a single rule or set of rules to identify referring expressions within chat. Rather than developing specialized algorithms to deal with the nuances of three-way, live chat in the VesselWorld domain, it would vastly simplify our task if users were to enter all the information the system needs in a common, structured form. Although this might seem like it would unnecessarily burden the user, below we will see that this is not the case.

4 Using CRs to fix representational problems

VesselWorld has representational deficiencies from the both the user's and the adaptive component designer's perspectives. In mediating the interaction between the users and domain, it does not offer the right kind of support for managing information about wastes. In mediating the interaction between the designer and the users, it does not provide structured access to the runtime information required for successful intent inference. The analyses performed above led to the design of a Coordinating Representation, which is a solution to both of these problems.

There are other approaches to designing of collaborative artifacts that address various issues in collaboration. Cognitive work analysis (Vicente 1999) advocates a host of ethnographic methods for developing interfaces that address collaborating users' needs. Activity theorists focus on the need for mediating artifacts that address tensions in the larger activity system (e.g. Engeström 2000). Suthers (2003) approaches the design problem from the perspective of the teacher, as a way to improve the process of collaborative inquiry. However, no existing framework draws together design problems from the perspectives of both the user and the adaptive component designer.

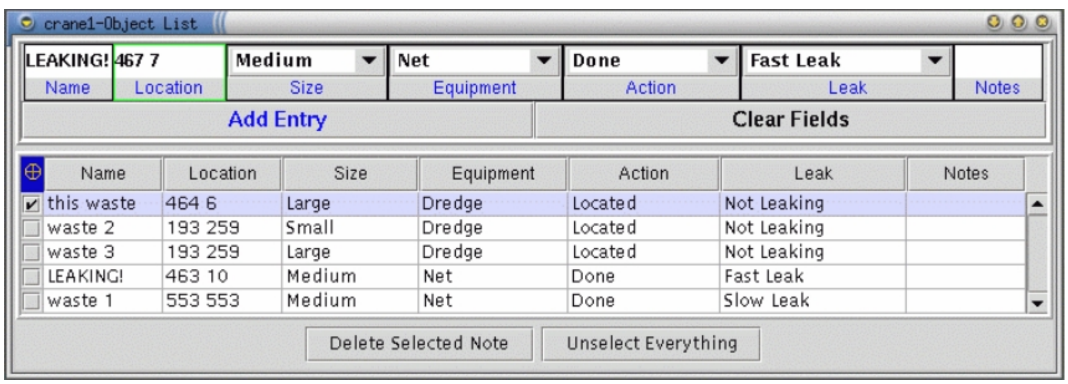


Figure 10: The Object List

Several CRs were developed for VesselWorld (Alterman et al. 2001). The CR that addresses the representational problems described above is called the Object List (Figure 10). The Object List is a WYSIWIS (What You See Is What I See) component that helps users to manage domain information and coordinate references. It displays the same data in tabular format for each user. Users enter and maintain all of the data in the Object List. Each row of data contains several fields of information, including a user assigned name, the status, and the location of the associated object. The location field may be filled in by clicking first on the field and then on the object on the map. The size, equipment, action, and leak fields are filled in using drop-down menus. A free text field ("Notes") is also provided for each entry so that any other relevant information may be communicated. Entries in the Object List can be displayed on the World View (Figure 5) as icons that are annotated with the name that is in the "Name" field.

There is nothing exotic in the design of the Object List. It is remarkable as the product of a repeatable design methodology that is explicit guidance for the creation of shared artifacts which simplify coordination for users and make intent inference easier. In the following, empirical data is provided that demonstrates the Object List indeed achieves both of these goals.

4.1 Improvements from the user's perspective

In all user experiments, the Object List was heavily used, and the coordination problems described above were no longer observed. The Object List had significant qualitative impacts upon the way users coordinated information about toxic waste barrels. Figure 11 compares sample dialogues from users who did not have access to the Object List (the left column) to those that did (the right column). It is immediately apparent that the group that had the Object List spent far less time talking about the details of each waste. None of their discussion appears to be designed to exchange information about the wastes per se, but rather is focused on strategizing.

Without the Object List	With the Object List
1. Crane2: what was small at 275, 400? 2. Tug1: sX 3. Crane2: ahve sX at 450 above that 4. Tug1: mD at 400 350 5. Tug1: yes, there is an sX at 275 250 as well 6. Crane2: I have two amall at 275, 400 and 275, 450 are these the same? 7. Tug1: no, there are two sX there 8. Tug1: well, there are actually three in almost a line from n-s 9. Crane2: at 400 and 450? what about the 275, 250? 10. Crane2: ok, so the southern exists? 11. Crane2: I'm almost there, nothing at 275, 250 12. Tug1: 300 350, 250 400, and 275 250 are my markers 13. Tug1: argh 14. Tug1: I mean 275 450 15. Crane2: ok, those sound good 16. Tug1: i don't know why I kee doing that.	1. Crane1: I got an XL! 2. Tug1: I got nothing, you luck basrstartd. 3. Crane2: I got an Xl and an L, mommy! ;) 4. Tug1: Merry christmas kids…. 5. Crane1: I'll map North third? 6. Tug1: I'll take middle 3rd. 7. Crane2: I'm at south-central. Tug, where are you? 8. Tug1: I'm jus nw of the barge, let me put that on the map… 9. Tug1: actually more w than n. 10. Crane2: With the LB in the corner, maybe our best bet is moving the SB NW and loading it with all the NW corner's goodies, which CRANE1: can map 11. Crane1: not a bad plan… 12. Tug1: Ok, I'll make a bit of a sweep around here while CRANE1: looks around. 13. Crane1: Tug, can you pick up the SB at your earlier opp? 14. Tug1: CRANE2: can map up on the way?

Figure 11: Dialogue from users before an after the introduction of the object list

Figure 12 depicts the information users entered into the Object List during the dialog in the right hand column of Figure 11. This demonstrates that the information previously exchanged via the chat tool has now been offloaded into the Object List.

⊕	Name	Location	Size	Equip	Action	Leak	Notes
☒	G1	556 465	XLarge	Unknown	Located	Not Leaking	
☒	A1	186 107	XLarge	Unknown	Located	Not Leaking	
☒	m	550 447	Small	None	Located	Not Leaking	
☒	A2	249 21	Large	Unknown	Located	Not Leaking	
☒	m	250 149	Small	None	Located	Not Leaking	
☒	m	449 349	Small	None	Located	Not Leaking	
☒	SB	305 310	XLarge	None	Located	Not Leaking	

Figure 12: Object List created during dialogue in Figure 8

To assess the impact of coordinating representations quantitatively, a formal study was performed comparing the performance of teams with and without CRs. Each test group consisted of three teams of three subjects. Subjects were a mix of area professionals, mostly in computer-related industries, and undergraduate students; all were paid a flat fee for the experiment.

The results of this study, shown in Table 1, compare the final five hours of play for each group (by which time performance had stabilized), and these results were normalized over a general measure of complexity for each of the problems solved. The performance of the test group that had the CRs was better across several measures: amount of chat, number of errors committed, number of system events generated (an indicator of interface work), and clock time.

Indicator	Improvement (reduction)
Communication	57% ($p<0.01$)
Domain Errors	61% ($p<0.2$)
System Events	38% ($p<0.06$)
Clock time	49% ($p<0.01$)

Table 1: Improvement of CR groups over non-CR groups; final 5 hours of play

The most significant effect is the 57% reduction in communication generated. This confirms the observation that a significant amount of communication was offloaded into the CRs. Also highly significant is the 49% reduction in clock time. Only slightly less significant is the reduction in system events (mouse clicks, etc.), down 38%. Additionally, overall domain errors (errors in performing domain actions which led to a toxic spill) were reduced by 61%. The variance of this measure was quite high due to the overall low frequency of this kind of error; this reduced its confidence below statistical significance ($p<0.2$).

The above results demonstrate that the introduction of the Object List modifies the representational system from the user's perspective in ways that improve the ability of the collaborators to coordinate information. It is shown in the following how the Object List also addresses the representational needs of the designer.

4.2 Improvements from the adaptive component designer's perspective

The Object List captures information about the world that would otherwise be exchanged in chat, and transforms it into a form that can be easily used in the intent inference process described above. In addition to information about where the wastes are, what equipment they need, and what size they are, the Object List also provides the system with a set of user-assigned labels. We observed that participants used these labels in chat, and this in turn led us to examine whether or not these references were predictive of domain intention.

Table 2 depicts the likelihood that a reference for an object will appear in chat during the three consecutive five minute windows prior to a lift of that object at time *t*. In the table, "Joint" and "Single" refer to whether a waste requires both or just one crane operator to lift. In the *~Lift* conditions, values reflect the likelihood some waste is referred to prior to a lift of a different waste.

	t-5 to *t*		*t-10* to *t-5*		t-15 to t-10	
	Joint	Single	Joint	Single	Joint	Single
Lift	.62	.42	.27	.15	.25	.08
~Lift	.15	.11	.10	.07	.08	.04

Table 2: Probability of reference preceding a lift at time t

As shown in the table, there is about a sixty percent chance that waste will be referred to in chat in the five minutes preceding the lift if that waste requires assistance, and about a forty percent chance if that waste can be lifted singly. Thus, chat references are predictive of lift actions for roughly a fifteen-minute window of time preceding a lift. On the basis of this analysis, we expanded our BN to include three five minute windows of chat history, with one node for each five minute window. The expanded network is shown in Figure 13.

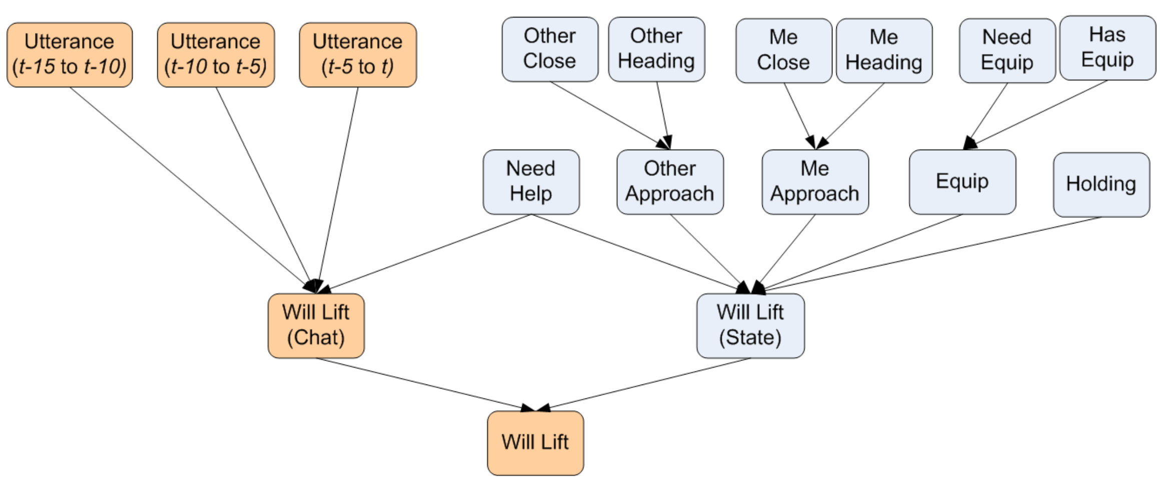


Figure 13: BN with nodes for chat added; added nodes are colored differently

The relative utility of information from the Object List for performing intent inference is evaluated in the following section. This is done by comparing the ability of the BN to predict user lift actions in historical log files across several information conditions. Before presenting results, the evaluation methodology is described in detail.

4.2.1 Evaluation Methodology

To evaluate the utility of information from the Object List, the BN was trained and tested under four information conditions using a single data set (described in Table 3), and its performance was compared across these conditions. The data set consists of time-stamped information derived from the log files collected from each use session in the data set. It contains the state information required by the BN. State information includes information about each waste (position, size, and equipment needs), the users' locations and heading, and wastes as they are lifted. The data set also contains information about references made by users during chat to objects that are listed in the Object List.

Team	Sessions	Avg. # of wastes per problem	Total Hours
Team 1	10	11.7	9.9
Team 2	6	11	8.4
Team 3	9	14.3	9.1
Team 4	16	14.5	8.7
All	41	13.5	34.3

Table 3: Data set for the evaluations

Each condition changes the amount and quality of information about wastes in the data set. The four information conditions, and the information available under each condition, were:

1. **Perfect Information** – Under this condition, the data set contains complete and

accurate information about the position, size, and equipment needs for every waste at each point in time. Perfect information is not typically available at runtime, but it can be derived from the problem files which define the initial problem configuration for a usage session. This is an ideal case, and provides a baseline against which to compare the other information conditions.

2. **Object List** – Under this condition, the data set contains information about the position, size, and equipment needs for wastes that are entered in the Object List, as it becomes available. This information is subject to errors, duplication, and omissions. Compared to the Perfect Information condition, performance of the BN under this condition is indicative of the quality of the information users provide via the Object List.

3. **Perfect info + Chat** – All of the information in the Perfect Information condition, plus the occurrence of names of objects (as entered by users into the Object List) in chat. More precisely, there is an entry in the data set for each time a word appears in chat matches one of the labels for a waste in the Object List (see Appendix A). This condition is used to investigate how much predictive information references to wastes in chat add.

4. **Object List + Chat** – Information from the Object List condition, plus the occurrence of names for objects in chat. This condition represents the best possible runtime performance of our inference procedure without information from problem files.

Because the data set contains different information under each condition, the probability distribution stored in the BN may be optimal for one information condition but not another. If the evaluation were to be performed using a single probability distribution for all data conditions, results might reflect the "preference" of that particular probability distribution, rather than the utility of information contained in the data set.

To overcome this problem, the probability distribution of the BN was fit to the data set in each information condition before testing. To train the network, each data set is compiled into a series of cases. A case consists of values for each non-hidden node in the BN (see Section 2.1), and each case represents a change in the relevant information about the world. During the training phase, values for the "Will Lift" node were extended back in the data set from each actual lift for a ten minute time window. The size of the time window was based on initial experimentation, and chosen to maximize the performance of the BN.

The *EM(η)* algorithm (Bauer, Koller, and Singer 1997) was used to train the network, and the same starting parameters for the BN were used in each case. The *EM(η)* algorithm is a generalized version of the standard *EM* algorithm with a learning parameter. When $1.0 < \eta < 2.0$, the *EM(η)* algorithm is significantly faster than the standard *EM* algorithm, but still has good convergence properties. For all of our training sessions, $\eta = 1.8$. Algorithm performance was further optimized by posting only unique cases to the network, and weighting each unique case in the training set according to the number of times it occurred.

For the evaluation, two performance metrics were calculated: the correct goal rate (CG_R), which is the proportion of correctly guessed goals; and the false positive rate (FP_R), which is the proportion of guesses that were false. A guess is made whenever a relevant state variable changes. Any uninterrupted sequence of correct guesses leading up to the step immediately preceding the execution of the predicted goal is counted as a single correct goal. The total number of goals is the number of wastes lifted. Thus,

$$CG_R = correct\ goals / total\ goals$$

$$FP_R = incorrect\ guesses / total\ guesses$$

4.2.2 Evaluation

The evaluation was performed solely to compare the utility of different information sources for intent inference, and in particular, the utility of the information derived from the object list. These performance metrics are not indicative of how well the intent inference procedure performs within the context of a specific adaptation. Studies of the adaptive component are documented in the next section.

The results of the evaluation are shown in Table 4, and reflect average performance across all teams for each data condition. A single factor ANOVA demonstrated that differences between information conditions were highly significant for both CG_R ($F(131,3)=10.84$, $p<.0001$) and FP_R ($F(131,3)=3.98$, $p<.01$).

Condition	Correct Goal Rate (StdDev)	False Positive Rate (StdDev)
Perfect Info	.83 (.14)	.53 (.13)
Object List	.70 (.17)	.60 (.16)
Perfect Info + Chat	.87 (.12)	.51 (.11)
Object List + Chat	.77 (.15)	.58 (.15)

Table 4: Intent inference results for different information sources

The “Perfect Info” case, in the top row of the table, provides a baseline against which results for the other conditions may be compared. It is a rough indicator of the best the intent inference procedure can do, given only complete information about the state of the world. Across the four teams in the dataset, the Correct Goal Rate for the "Perfect Info" case ranged from .77 to .91 for this condition. There was a weak correlation between problem size (number of toxic wastes) and individual team performance ($r=.23$), reflecting the fact that it is more difficult to make good guesses when there are more options to choose from. In general, these metrics indicate that the inference procedure is effective.

As expected, the intent inference procedure does not perform as well with information from the Object List alone. However, results from the "Object List" condition were still good, and demonstrate that use of the Object List was reliable enough to be useful for intent inference.

The "Perfect Info + Chat" condition demonstrates that references add significant information that cannot be derived from knowledge about the state of the domain. Thus, regardless of access to state information (for instance, if there were intelligent sensors placed in the world) the Object List adds information that still improves intent inference.

The combination of reference information from chat and domain information from the Object List (the "Object List + Chat" condition) improves the performance of the procedure to a point where it is nearly as good as the "Perfect Info" condition. These results confirm that, by modifying the representational properties of the media between the user and adaptive component designer, high quality runtime information can be made available for intent inference.

To demonstrate that this information is sufficient to drive a useful adaptive support mechanism at runtime, an adaptive component was implemented and its use evaluated. These results are described in the following section.

5 An Adaptive Component

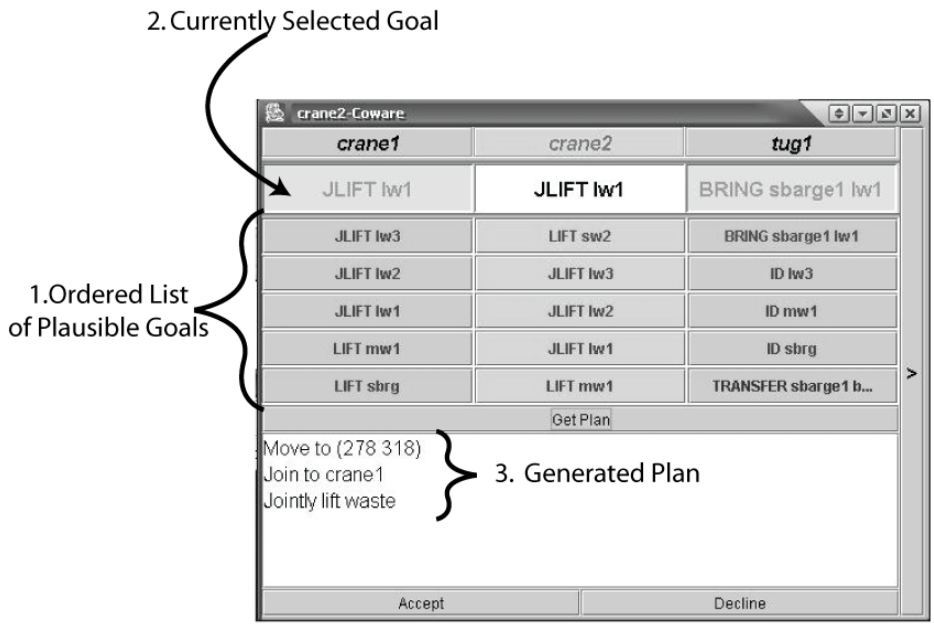


Figure 14: The adaptive component

As described in Section 2.1, planning in VesselWorld is a laborious and error prone operation, and user errors are frequently the result of forgotten plan steps or joint plans that have become unsynchronized. The adaptive component was thus designed to help users formulate basic individual and joint plans. Using the intent inference procedure, the system can provide a small set of likely plans from the hundred or so that are possible at any point in time.

The interface to the component is shown in Figure 14. It displays inferred plausible domain goals for each of the users. Like the Object List, it is a WYSIWIS component, so each user can see the other users' goals as well. Users can select plausible goals from their own column to have the system automatically generate a plan. The component can generate individual as well as shared plans. If the information in the Object List is correct, the generated plans are guaranteed to be correct.

The detailed function of the adaptive component is as follows:

1. Each time state information is updated (e.g., when plans are executed, information is added to the Object List, an object reference appears in chat, etc.) the system offers each user up to five plausible goals, displayed in order of priority and color-coded according to the system's "confidence" in that prediction. Each user can only select goals from their column.

2. When the user selects a goal, it is copied into the top row, which displays each user's currently confirmed goal. The user then has the option to request an automatically generated plan that will accomplish the selected goal (the "Get Plan" button shown in Figure 14).

3. The system generates a plan that the user can inspect. In cases where the goal involves multiple actors, the other actors are invited to join the plan. If all invited actors accept the invitation, a plan is generated; if invited actors do not accept the invitation, the requesting user is so informed.

4. If the user accepts the plan (by clicking the "Accept" button in Figure 14), it is automatically copied into the user's planning window for execution.

If the plan is generated from correct state information, no user modifies the state in such a way that conflicts with the generated plan, and the plan is executed to completion, it will succeed in achieving the desired goal.

5.1 User Studies

We performed a formal study with four teams to evaluate the use of the adaptive component and its effects on performance. Two of the teams were given access to the adaptive component, and the others were not. The participants were a mix of students and local-area professionals, with varying degrees of computer proficiency. Each team was trained for two hours in use of the system, and then solved randomly chosen VesselWorld problems for approximately ten hours. To alleviate fatigue concerns, the experiment was split into four three-hour sessions.

For the teams with the component, the inference procedure used information from the Object List and chat (the Object List + Chat condition) to infer user goals and construct plans. The following results report on the last 5 hours of play time for each group, by which time performance of the users had stabilized.

5.1.1 The component was heavily used

All groups used the component to generate plans within the system. On average, users confirmed a goal every 1.5 minutes (SD=46 seconds), requested a plan for each confirmed goal, accepted 71% of plans requested (SD=19%), and completed the execution of 83% (SD=6.75%) of these plans. Overall, this indicates that roughly 59% of confirmed goals resulted in a plan that was executed to completion. For each problem solving session, one quarter of all plan steps submitted to the server were generated by the component (SD=8%).

The component generated plans for 43% (SD=15%) of the domain goals it could have predicted for the Cranes (some goals, like "search the harbor" were not inferred). It was not possible to obtain a similar statistic for the Tug operator because it is difficult to recognize the tug's goals in the collected log files (goals for the tug are not bracketed by easy to detect plan steps like "LIFT" and "LOAD").

5.1.2 The component improved users' performance in the domain task

The groups that had the component had 45% ($p<.10$) fewer joint errors (failures during joint actions) per minute than the groups that did not. This difference is not significant at the .05 level because of the small sample size and overall low proportion of joint errors. A reduction in joint errors corroborates prior analysis of use of the VesselWorld system, which indicated that joint errors were usually the result of plan submissions becoming unsynchronized (Alterman et al. 2001). Because the component generates coordinated plans in advance, users could simply submit each step and be assured that actions would be coordinated.

5.1.3 The component reduced cognitive effort

To measure the change in cognitive effort between the two populations, the amount of time it took users to execute plans and the amount of interface work were evaluated. It was found that the amount of clock time taken by users between submitting steps of automatically generated plans was 57% less ($p<.01$) than in groups without the adaptive component, but that there were no significant differences in the number of mouse clicks per waste. Because the reduction in clock time for groups with the component cannot be explained by a reduction in the amount of interface work, we conclude that the component reduced the cognitive effort of the collaborators.

As stated in Section 1.2, these studies show that the adaptive component was heavily used, improved user performance, and reduced cognitive effort during plan execution. This is verification that collaborating users can generate enough structured information when using a coordinating representation to drive useful intelligent support. We conclude that the approach to adding intelligent support demonstrated here was successful for this domain.

6 Developing Novel CRs

This work has demonstrated how the design of a groupware system can be modified to address the needs of two intersecting representational systems. This is done by adding structure to the system in the form of a Coordinating Representation. The CR helps users coordinate and improves performance in the domain task. It also renders information previously exchanged via an unstructured communication channel accessible for use in an adaptive support system.

Clearly, much in this approach rests on the design of the CR. In some domains, CRs have already been developed and explored by other researchers. However, in novel domains such as VesselWorld, a methodology is necessary to guide the development of new CRs. Such a methodology has been developed in work that supports the work presented here (Alterman et al. 2001; Landsman and Alterman 2005; Feinman and Alterman 2003).

This methodology can be summarized in four steps:

1. Transcripts are collected of runtime user behavior.
2. These transcripts are analyzed in order to identify weak spots in the user's representational system.
3. This analysis is used to develop coordinating representations that people will use, and that improve coordination.
4. Information collected by the CRs is leveraged to drive adaptive components.

Transcript collection has roots in ethnography and interaction analysis. Traditionally, ethnography has relied upon *in situ* observation. Suchman and Trigg (1992) and Goodwin and Goodwin (1996) explored the use of video recordings in capturing and analyzing naturally occurring workplace activity. While video is a very powerful technique for studying a work practice, it is a resource intensive process. For the ethnographer, groupware makes life much easier, because a large portion of the relevant activity is mediated by the system itself, and can thus be passively recorded and analyzed offline.

To perform the kind of ethnographic analysis required to identify representational deficiencies in groupware, it is helpful to be able to replay transcripts, rewind when necessary, easily locate high-level task events (e.g., the submission of plans), and add annotations. To provide this functionality, a component-based software framework called THYME was developed (Landsman and Alterman 2005). Groupware systems built using THYME automatically log all usage data, and replay tools can be rapidly generated. The replay tools provide a console similar to that on a VCR, but with all of the above features.

Two techniques have been developed to analyze transcripts. One of them, which was described briefly in Section 3.1, examines users' recurrent activities to identify weak spots in an existing representational system (Alterman et al. 2001). The other technique examines co-reference chains (Feinman and Alterman 2003). To perform this later analysis, the analyst identifies and tags all references to objects, plans, and other entities of interest in a logged dialogue. Metrics are then calculated for the distribution and lifespan of each type of reference and these metrics are used to identify the need for supporting structure. Empirical evidence has been collected demonstrating that these analysis techniques can be used to make predictions about the type of structure that will be useful (Feinman 2006).

These analyses guide the development of CRs that address representational needs of both the user and the designer. Because CRs fix problems with an existing representational system for the users, they will improve the users' ability to stay coordinated, and users will use them. This chapter has demonstrated how the structure introduced by CRs can than be leveraged to introduce adaptive support.

7 Summary

As this chapter has illustrated, coordinating representations offer a means for merging two traditionally competing design requirements for adaptive systems into a single representational medium. On the one hand, CRs can reduce user workload and improve coordination. On the other, by structuring coordinating information, CRs provide the designer with a rich source of user and context information that can be leveraged by an adaptive support system at runtime.

Thus, CRs are a viable knowledge acquisition strategy in groupware. In systems where CRs have not yet been identified, such as the one presented here, there is an analytical procedure that supports their development. This methodology considers coordination problems to be an opportunity for the development of useful, structured mediation, and it can be applied to any media that a community of actors uses to coordinate a recurring activity. Relatively complete usage data is required to perform the prescribed analysis procedure. In the case of groupware that mediates the majority of the users' interactions, this may be achieved through the addition of logging utilities to the groupware system itself. For collaborations which extend outside of the media under investigation, more traditional forms of data collection (e.g. videotape) will be required.

The approach presented as been validated in a groupware platform designed to mediate interaction wih a simulated domain called VesselWorld. Nonetheless, the coordination problems experienced by VesselWorld users are highly representative of those problems in other collaborative environments. Users have different information about the world and need to share this information to accomplish their joint task. Individuals change objects in the world and need to inform others of these changes. Users have different roles and responsibilities. The task requires close coordination at times, but long term commitments are also a critical part of the activity.

In this chapter, we've focused primarily upon problems related to grounding, but our analysis has revealed other types of coordination problems as well. Alterman et al. (2001) document the problems users faced in sequencing closely coordinated activities and tracking long-term commitments, and describe CRs that were developed to address these problems.

Similar kinds of problems have been described in the context of modern military applications. As networked technology has become ubiquitous in the military, electronic chat has found its way into mission-critical applications. Usage patterns reminiscent of the coordination problems in VesselWorld have been documented. Chat history is heavily relied upon to recall missed or forgotten information, and there are problems keeping track of mission orders (Heacox et al. 2004). The need for adaptive chat mechanisms that are more aware of the user's task so as to minimize interruptions (Cummings 2004) has also been noted.

Because CRs produce domain relevant, structured information, very little effort needs to go into converting that information into a usable form. This enables the use of common "off-the-shelf" inference techniques. In the case study presented here, Bayesian Networks were employed because they are easy to use, were sufficiently powerful for our needs, and many open-source and off-the-shelf implementations are available. Other techniques might have been used instead of or in addition to BNs. Sophisticated NLP algorithms could likely derive more mileage from references to wastes in chat than we were able to with our procedure. The training algorithms used to analyze the BN's performance could be employed at runtime to adapt to user behavior over time. More sophisticated planning algorithms and optimization routines could have been employed to produce efficient plans given available waste information. The point is that a range of possibilities for building adaptive support cascade from the increased availability of structured information about the user. The approach is then limited by the relevance of the information to the adaptation goal, the power of available inference algorithms and expertise of the designer.

The end result of applying our methodology to VesselWorld was a successful adaptive plan-generation component. This component is entirely domain specific, and may not be necessary or feasible in other domains. A natural extension to the existing methodology would offer guidance to the designer regarding the type of adaptive support that may be desirable or feasible using coordinating information from CRs.

One possibility is to use the information extracted from the CR to further mediate the users' communication in order to improve it with respect to some ideal. As described previously, Goodman et al. (2005) discuss how a group's discussion can be monitored for instances of poor collaboration, and subsequently modified via the injection of adaptive support. One drawback of the approach they described was that additional non-domain work was required on the part of the users in order to provide the right kind of information for the adaptive component. The strategy presented in this chapter is a way to address that drawback.

In the following chapters I will describe an adaptive system that is similar to the approach described by Goodman et al. (2005), in that I seek to influence sub-optimal collaboration (with respect to an abstract ideal) via an adaptive component. However, the information I use is made available from a CR that is designed to help users in their domain task, and the specific approach I take to mediating the collaborative process is somewhat different.

The next chapter covers the relevant background necessary to ground my second case study. The domain under investigation is group decision making and covers a broad range of topics, including decision analysis, group information processing, argument visualization and group decision support systems. Following this background material I will briefly introduce the design of an adaptive platform that builds upon an argument visualization based CR that borrows heavily from prior work. The remaining chapters will describe a case study, and present a detailed analysis of the data collected in that study.

Chapter 3: Adaptive Mediation in Group Decision Making

The first two chapters of this dissertation discussed two opportunities that may be exploited when developing adaptive functionality for a given groupware platform, both of which exist because groupware mediates a group's communication about their task. The first opportunity concerns knowledge acquisition. The preceding chapter illustrated how a mediating structure, called a coordinating representation (CR), can be used to take advantage of this opportunity and make quality information available at runtime to drive the adaptive algorithm.

The second opportunity concerns how this information might be used once it has been acquired. Because a groupware system mediates all collaborative interaction, the designer has an opportunity to introduce adaptive support that can transform the collaborative process to overcome dysfunction or suboptimal performance. I have previously referred to this type of adaptation as *adaptive mediation*. To develop this type of adaptive system, the designer requires a theory (or several) about what the collaborative process should look like. Thus, the adaptive functionality is not incorporated simply to enable the system to respond to changing usage scenarios; it is incorporated to transform the user's activity into something more like an ideal in the mind of the designer. A careful specification of this ideal should precede the design of adaptive mediation in collaborative systems.

In this chapter I will lay the ground work for a system that is able to overcome dysfunction in group decision making. In developing a system of this nature, three questions will need to be answered:

1. What is the problem with the collaborative process that we wish to address?
2. For the problem we've identified, what is the ideal collaborative process?
3. What type of CR can we use to solve the knowledge acquisition problem?

Accordingly, the following review is divided into three components, which will provide answers to these questions. First, will describe a problem with group decision making that my platform will address. This problem has been referred to as the "common knowledge" problem, and describes decision making groups' tendencies to focus on information everyone in the group shares. As I will describe, some efforts have been made to address the common knowledge problem with group decision support platforms, but these efforts have had mixed results. These negative results clarify the problem my platform seeks to address.

I'll then turn to some of the models that might guide me in developing adaptive mediation for a group decision making process. This review will focus upon models that are based upon, or have been part of, decision analytical frameworks that derive from Savage's (1954) Subjective Expected Utility (SEU) theory. I will also cover some of the group decision support tools that have been based on these models, and discuss their drawbacks. Note that there are other models of decision making, but discussion of these will be postponed until later chapters, when this material becomes relevant.

Following this analysis of rationalistic models and the the tools that are based on them, I will turn to research on argumentation and argument visualization in order to identify a CR that can support deliberative dialog. After covering this literature, I briefly review a couple of systems that have combined argumentation with SEU-style decision analysis. I will then offer a brief summary to clarify the answers to the above questions, and describe the skeleton of an adaptive groupware platform designed to overcome the common knowledge problem.

1 Problems in Group Decision Making

Critical decisions are often made by groups. The most common reason given for this is that groups can access a much larger pool of information and expertise than individuals alone (e.g. Shaw 1981). However, information sharing in groups is imperfect, leading groups to make poor decisions despite their potential advantages (c.f. Janis 1982; Myers and Lamm 1976; Stasser and Titus 1985). In the following, I describe some of the general theory related to group decision making, and then refine my focus to a particular problem that is the focus of the subsequent chapters.

Group decision making has been characterized as consisting of three activities; information retrieval, information exchange, and information processing (Dennis 1996; Briggs 1994). Information processing may further be broken down into two potential routes to preference change – one being central, which involves the careful consideration of information, and the other being peripheral, which describes processing that depends on factors other than the information itself (Petty and Cacioppo 1986). The relationship between these activities is shown in Figure 15.

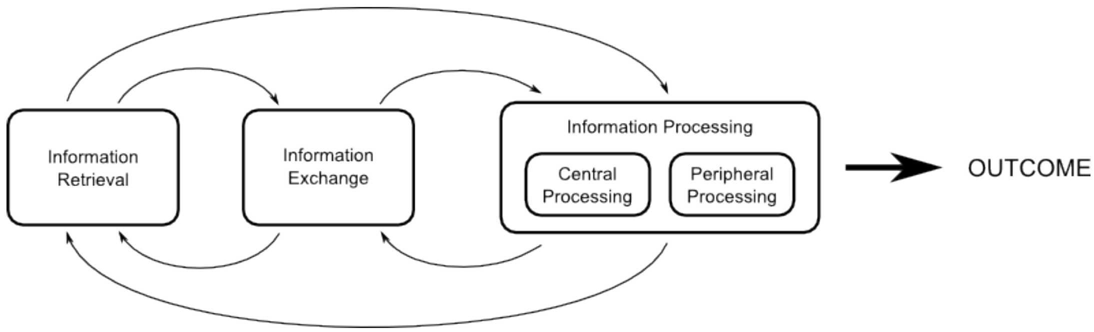


Figure 15: Notional group decision making process

They do not occur in order, but individuals are generally only able to engage in one of these activities at a time, and they compete with each other for cognitive resources (Norman 1976). In face-to-face situations, the competition between these various types of cognitive activities can lead to a number of inefficiencies in collaborative information processing. These inefficiencies, and some of the other factors that affect each of the above three phases are described in the following.

Information Retrieval. Information can be retrieved either from memory, or external resources. Information retrieval in face-to-face discussions can suffer because participants must concentrate on understanding information that is being discussed (Lamm and Trommsdorff 1973). Information retrieval also tends to be biased towards that which supports an individual's point of view (Petty and Cacioppo 1986), though the availability of external information resources can mitigate these tendencies (Hollingshead 1996).

Information Exchange. Communication plays a central role in group decision making.

In group discussions, there are a number of factors which enhance or impair information exchange; these can be characterized as either process gains and process losses (Steiner 1972). Table 2, adapted from Nunamaker et al. (1991) summarizes some of the potential process gains and losses that might occur as a because information is exchanged in a group-setting.

Common Process Gains	
More information	A group as a whole has more information than any one member.
Stimulation	Working as a part of a group may stimulate and encourage individuals to propose new ideas, or retrieve other relevant information.
Common Process Losses	
Air Time Fragmentation	The group must partition available speaking time among members.
Production Blocking	Information is not exchanged because of the competing cognitive demands of information retrieval, information exchange, and information processing; participants or are unable to find an opportunity to contribute and / or forget to exchange information. Several kinds of production blocking include attenuation blocking, concentration blocking, and attention blocking.
Failure to Remember	Members lack focus on communication, missing or forgetting the contributions of others.
Social pressures	Information is not exchanged because participants fear negative evaluations, reprisals, or are simply trying to be polite.
Cognitive Inertia	Discussion moves along one train of thought without deviating because group members refrain from contributing comments that are not directly related to the current discussion.
Domination	Some group members exercise undue influence or monopolize the group's time in an unproductive manner.
Information Overload	Information is presented faster than it can be processed.
Coordination Problems	Difficulty integrating members' contributions because the group does not have an appropriate strategy, which can lead to dysfunctional cycling or incomplete discussions resulting in premature decisions.

Table 5: Selected process losses and gains; adapted from Nunamaker et al. 1991 [1]

Some of the common process losses shown in Table 5 work together to cause group

1 I have taken a couple of liberties with the phrasing for summarization purposes; Nunamaker breaks “production blocking” into three subcategories, and “social pressures” into two categories. However, the semantic content of the categories that are summarized has not changed. I have also omitted some categories of process losses and gains that are not relevant to the current investigation. Readers are referred to the original source for the complete list as well as references to primary sources of these process losses and gains.

discussion to focus on information that most of the group members hold in common prior to group discussion. Social pressures may lead people to be less likely to contribute opinions that do not conform to the majority opinion, and both production blocking and cognitive inertia tend to focus group discussion on the same confirmatory subset of information. This can lead to group polarization (Myers and Lamm 1976) and groupthink (Janis 1982).

Information Processing. Information processing refers to the synthesis of exchanged information. As mentioned above, and shown in Figure 15, two theoretical routes have been used to describe how the information exchanged by a group is processed (Petty and Cacioppo 1986; see also Winquist and Larson Jr. 1998). The "central" route describes how participants actively evaluate information and integrate it into their overall understanding. Along the "peripheral" route, participants' opinions are shaped by other cues that require less cognitive processing, such as the majority point of view, or other social ad contextual factors (e.g. the tone or setting of an interaction).

One theory that explains how information is processed along the central processing route is the *persuasive arguments* theory (Vinokur, Trope, and Burnstein 1975; Burnstein and Vinokur 1973). The key idea in the theory is that it is the information itself which matters, and not the preferences of others with respect to that information. The theory predicts that novel information should have more of an impact on participant decision making than information that is already known. However, this is at odds with research that has demonstrated that people in groups focus more upon information they hold in common (Stasser, L. A. Taylor, and Hanna 1989; Gigone and Hastie 1993). The importance of novel information can be artificially raised when participants are encouraged to actively process incoming information (Postmes, Spears, and Cihangir 2001). However, several studies have shown that generally, people process the information that supports their position more thoroughly (Petty and Cacioppo 1984) and actively develop counter arguments to information that does not (Wood 1982; Lord, L. Ross, and Lepper 1979).

Unlike central processing, peripheral processing in groups tends to rely upon information that is easier to use. *Normative influence* theory suggests that the dominant form of peripheral processing is rooted in interpersonal comparisons (Myers and Lamm 1976). Normative influence theory predicts that exposure to others' preferences can cause an individual to modify their own preference to agree with the majority (Hackman and Kaplan 1974), or simply accede to the majority preference without any change in their true preference (Maass and R. D. Clark 1984). Social judgment theory (Brehmer and Joyce 1988) and social decision schemes (Davis 1973) offer compatible explanations for peripheral processing (Winquist and Larson Jr. 1998). Such theories support the view that an individual's final judgment can be described mathematically as a combination of all individual preferences, entirely bypassing central processing.

Neither central nor peripheral processing routes entirely explain observed group decision making behavior (Zuber, Crott, and Werner 1992; Myers and Lamm 1976; Shaw 1981). Under controlled conditions, both kinds of processing can be observed, and within the body of research on group decision making, advocates can be found for either position. In the following section, I review a widely reported problem with group decision making.

The problem touches upon many of the above issues, and further clarifies the underlying debate between central and peripheral processing. That is, do people in groups function as a rational information processing collective, or are people fundamentally inclined to use other (non-informational) means in a social settings to make decisions?

1.1 The Common Knowledge Problem

A well-established and carefully studied problem with small group decision making is the inability of a group to pool and process all the information available to its constituent members. As discussed above, groups tend to focus their discussions preferentially upon information that members hold in common prior to group interaction. In addition, pre-discussion biases (biases that individuals have prior to group discussion) have substantial influence on decision outcomes, "as if group members exchanged and combined their opinions but paid little attention to anything else" (pg.132). This problem has been referred to as the "common knowledge" effect (Gigone and Hastie 1997). The primary vehicle for examining the common knowledge effect is the *hidden profile* experiment, first introduced by Stasser and Titus (1985).

In a hidden profile experiment, members of a group are asked to make a choice between several alternatives after pooling their information. Information is distributed among participants so that some individuals have some information that others don't, and the correct choice can only be determined by considering all of the information.

A typical formulation for the hidden profile experiment is portrayed in Figure 16 (adapted from Dennis 1996). In the figure, information for and against two options (A and B) is distributed among three collaborators. Information shared by the participants is referred to as "shared" information, and information held by only one of the participants is referred to as "unshared" information. By itself, the shared information recommends decision option "A" because there are two pieces of shared information in favor of A, and one piece against B. Each individual should also prefer option "A" because each individual has two pieces of information in favor of A, two pieces of information against B, and one piece of information against A. However, if all shared and unshared information is considered together, a decision option B should be favored. This is the typical structure of hidden profile studies, and it allows psychologists to study the effectiveness of information pooling in group decision making.

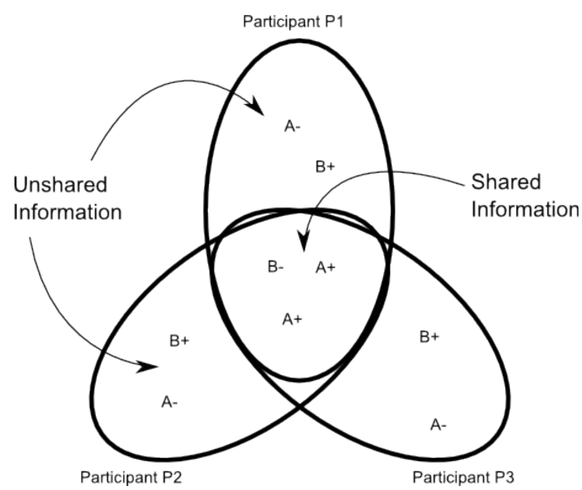


Figure 16: A hidden profile situation. Three participants share three pieces of information about candidates A and B, and each participant has two pieces of information not shared with others. The shared information recommends candidate A, but all information combined supports candidate B. Adapted from Dennis, 1996.

In their seminal paper, Stasser and Titus (1985) found that groups engaged in a hidden profile task were less likely to use unshared information and hence less likely to identify the hidden profile (the correct solution), a result seemingly at odds with predictions based on persuasive arguments theory. Instead, the group's decision was best predicted by the distribution of pre-discussion preferences, and consistent with shared information. Numerous manipulations of experimental variables and more precise analyses have elaborated upon this finding (see Wittenbaum, Hollingshead, and Botero 2004, Stasser and Titus 2003, and Kerr and Tindale 2004 for reviews). Shared information is more likely to be mentioned than unshared information during group discussion, and it is mentioned more frequently (Stasser and Titus 1987). It is also mentioned earlier and this can lead groups to premature consensus (Larson Jr. et al. 1996). Groups with less shared information and lower information loads discuss proportionally more unshared information, but decisions still get made on the basis of shared information and pre-discussion preferences (Stasser and Titus 1987). And even though shared information is discussed more and has more influence than unshared information, group judgments are best predicted by the majority of members' pre-discussion judgments (Gigone and Hastie 1993; Gigone and Hastie 1997).

Some manipulations have been shown to improve face-to-face group performance in hidden profile tasks. Disagreement among pre-discussion judgments leads to improved performance (Brodbeck et al. 2002). Groups are better at solving hidden profiles when they are told there is a correct solution (Stasser and Stewart 1992). Groups with critical norms – that is, those groups encouraged to engage in critical thought – have been found to perform better than those with consensus norms (Postmes, Spears, and Cihangir 2001). Asking members to rank-order alternatives rather than select the best option improves performance (Hollingshead 1996). Hidden profiles are solved more effectively when some of the members in a group are known experts (Stasser, Stewart, and Wittenbaum 1995). Reduced time pressure seems to facilitate the exchange of unshared information, and improve decision quality (Bowman and Wittenbaum 2002; see also Larson Jr. 1997). However, none of these manipulations completely eliminates the common knowledge effect.

There are currently a handful of theoretical mechanisms used to explain the phenomenon, and these theories are aligned with either the central or peripheral route information processing perspectives (Wittenbaum, Hollingshead, and Botero 2004). The first theory to be proposed suggested that probabilistic sampling is the main culprit (Stasser and Titus 1985). Simply, because more people have been exposed to the common information, it is more likely to be retrieved and discussed in conversation. Larson Jr. (1997) extended this explanation to include a temporal component, and his model has yielded fairly accurate predictions about the quantities and serial positions of unshared data exchanged during group discussion. Note that these theories do not conflict with the central route processing perspective. They place the cause of the problem with information exchange and the general fallibility of human memory. The group outcome reflects common information simply because the group bases their decision on a pool that is dominated by common information.

However, sampling theories do not explain why pre-discussion opinions matter so much in group decisions (Winquist and Larson Jr. 1998; Gigone and Hastie 1993; Gigone and Hastie 1997). An alternative explanation is that pre-discussion preferences influence group decision making by biasing information processing at the individual level. Greitemeyer and Schulz-Hardt (2003) showed that preference consistent information is evaluated more positively than preference inconsistent information in the context of a group decision making task. This is compatible with previous findings that people tend to process information they agree with more thoroughly (Petty and Cacioppo 1984) and develop counter arguments to information they disagree with (Wood 1982; Lord, L. Ross, and Lepper 1979). Because hidden profiles organize information so that preference consistent information is mostly shared and preference inconsistent information information is mostly unshared, hidden profile results can be explained as a direct result of individual level phenomena.

As with sampling theories, the above perspective is also consistent with the central route processing. The problem occurs at the information retrieval level (prior to information exchange) because group members evaluate information based on their individual credibility assignments. Once again, group decision making fails because of bad input, not because of bad information processing.

A third theoretical position is that the common information effect is a result of social comparison processes (Festinger 1975). Hearing that others possess the same information increases the importance and credibility of that information (Postmes, Spears, and Cihangir 2001; Wittenbaum, Hubbel, and Zuckerman 1999). Furthermore, social validation from others can lead group members to prefer repeating information that is known to be shared by others. This is consistent with findings that socially acknowledged role assignments can reduce hidden profile effects by increasing the credibility of hidden information even in the face of contradictory majority opinion (Stasser, Stewart, and Wittenbaum 1995). Unlike the previous two theories, this is a peripheral route explanation for the common information effect, and places the blame for the common knowledge phenomenon squarely upon group information processing, rather than retrieval or exchange.

There may be components of each of the above theories that are partially responsible for observed phenomena. It is true that group information pooling is biased towards common information, that pre-discussion bias explains much of the variance in the outcomes of hidden profile experiments, and that social comparison influences group dynamics. Some of the literature on group decision support systems sheds further light on the problem.

1.2 GDSS and the Common Knowledge Problem

Group Decision Support Systems (GDSSs) are groupware systems that are designed to support collaborative decision-making (DeSanctis and Gallupe 1987). In their most rudimentary form (so-called "Level 1" GDSSs), they contain support for asynchronous discussion via a message board and some sort of voting tool. These tools can reduce process losses by supporting parallel communication, improving group memory, and enabling increased anonymity.

These three aspects of GDSSs have been frequently identified as presenting an opportunity to overcome the common knowledge problem (Dennis 1996; Lam and Schaubroeck 2000). Parallel communication and group memory may reduce production blocking enough so that participants are better able to retrieve unique information.

Increased anonymity may reduce some of social pressures that increase normative information processing. Anonymity might also encourage minorities to persist in supporting their viewpoint, which leads better group processing of uniquely held information (Dennis, Hilmer, and N. J. Taylor 1997).

Numerous studies have been performed to examine whether indeed GDSSs can address the common knowledge problem, but these have yielded conflicting results. Hightower and Sayeed (1996) report that GDSSs lead to more biased discussions and poorer decision quality. Straus (1996) found that they have little discernible effect in hidden profile tasks. Lam and Schaubroeck (2000) report that GDSS improves the exchange of unshared information and decision quality in hidden profile situations. To complicate matters, many studies use slightly different experimental designs, making it difficult to synthesize results.

Despite these contradictory findings, there are a handful of studies that have consistently found that the features provided by a Level 1 GDSS can overcome some process-losses and improve information exchange (Dennis 1996; Dennis, Hilmer, and N. J. Taylor 1997; Mennecke 1997; Shirani 2006). In each of these cases though, decision quality is no better than groups without the GDSS, despite improved information exchange. Thus, even though groups produce enough information during the discussion to allow them to identify a hidden profile, they are still unable to do so.

Reflecting on the analysis in the previous section, this latter result indicates that at least one component of the common knowledge problem is indeed ineffective information processing; that is, based on studies of the common knowledge effect in GDSS settings, peripheral route information processing is partially responsible for the problem.

This is the key insight upon which I base my work. GDSS seems to be able to increase overall information throughput, but does not address dysfunctional information processing. My goal, then is to develop a GDSS system that is equipped with some adaptive functionality that will guide collaborators to process information along the central route.

In the following section, I will explore various models of decision making that might serve as a basis for my envisioned adaptive functionality.

2 Models of Decision Making

Abstract models of decision-making exist in part because what constitutes a "good" decision is quite hard to pin down. A great decision maker might make the wrong decision due to random events beyond their control. Conversely, a lousy decision maker might get lucky and make the right decision despite the absence of a good approach. For this reason, decision scientists usually distinguish between a "good decision" and a "good outcome." A good decision does not guarantee a good outcome, but should make good outcomes more likely. To figure out what a good decision looks like requires a model against which a given instance of decision making might be compared.

In the following, I will focus upon models that have originated from the "rationalist" school of decision making. It should be noted that the rationalist approach is not the only approach. There are many other rich troves of work which might be drawn upon when studying decision making, in particular some of the more recent work on naturalistic decision making (e.g. Klein, Orasanu, and Calderwood 1993), and work describing group processes (e.g. Poole and Doelger 1986). These will not be discussed here, because my platform is based upon the rationalist school. However, some of these approaches will become relevant in subsequent chapters, and will be discussed there.

Bell, Raiffa, and Tversky (1988) segregate decision making models into three categories: *normative* models, which describe what decision making should look like in an ideal case; *prescriptive* models, which suggest pragmatic strategies for moving the "what is" closer to "what should be;" and *descriptive* models, which describe decision making as it actually occurs. These categories provide a useful organizational framework, but they are not always easy to apply in practice (Bell, Raiffa, and Tversky 1988; pg. 30; see also Lipshitz and M. S. Cohen 2005).

The dominant normative perspective in decision analysis, and the cornerstone of the rationalist tradition, is Savage's (1954) subjective expected utility (SEU) theory. Savage's theory relies upon two fundamental assumptions. First, it assmes that an agent has a set of well-defined and consistent set of preferences or desires. By "consistent," it is meant that preferences obey the law of transitivity – if a decision maker prefers A to B and B to C, then the decision maker will also prefer A to C. The second assumption is that an agent should is able to cleanly disentangle the preference for an outcome from the likelihood of that outcome – that is, a decision maker's desire for an outcome should not influence the decision maker's assessment of the likelihood of that outcome. The SEU model asserts that if the choice (e.g., whether or not to finish my dissertation) of an option (e.g., I finish it) will result in a set of uncertain outcomes $\{x_i\}$ (e.g., x_1=I will get rich, x_2=I will be famous, and x_3=I will become tremendously athletic), the *expected utility* of that option can be determined as:

$$\sum_i u(x_i)P(x_i)$$

where $u(x_i)$ is the subjective utility of that outcome for the decider, and $P(x_i)$ is the subjective probability that outcome will occur. The rational agent will choose the option that results in the highest expected utility.

The SEU model forms the basis of modern economic theory, which assumes that in the aggregate, SEU is a reasonable approximation of human behavior (Bell, Raiffa, and Tversky 1988). Unfortunately, it has also been well established that the assumptions underlying the SEU model do not hold at the individual level. People do not maintain a consistent set of preferences for all possible outcomes, and it has been empirically demonstrated that preferences are both influenced by and influence expectations of likelihood (Kahneman and Tversky 1982; see also Cohen 1993). Thus, individuals violate both of Savage's fundamental assumptions. It has also been well established that people are very bad estimators of mathematical likelihood (Shafer 1988).

In the rationalist tradition, the goal of the analyst is to help guide people away from this apparent irrationality towards a normative ideal that is based on the SEU model. The way this is typically done is to decompose complex judgments into a set of simpler judgments that are easier to make, and then to recompose the results to determine the appropriate outcome (Shafer and Tversky 1988). Generally, an explicit mathematical approach is used to model the problem. In addition to suggesting a strategy for decomposition, mathematical approaches also help people estimate probabilities or express a consistent set of beliefs. Few approaches actually support both of these processes equally well. In the following sections, I will discuss several such approaches.

2.1 Likelihood models

A likelihood model dictates a probability semantics, and these semantics in turn suggest different types of problem decompositions. Probability semantics provide people with a metaphor for mapping their own internal judgments of likelihood into the domain of mathematical probability. The kinds of analogies they invoke to help people do this suggest different sorts of variables to be modeled. Which approach is chosen depends on the type of problem to be modeled, and the knowledge and experience of the modelers. In the following, I describe two dominant approaches, and modeling frameworks that are compatible with these approaches.

2.1.1 The Bayesian Approach

In the Bayesian approach, probabilities are given the standard mathematical interpretation – that is, in a chance experiment, $P(A)=p$ indicates that outcome A will occur approximately p percent of the time given a large enough sample. To help the modeler express his own internal representation of likelihood as a mathematical probability, one of several interpretations of probability may be invoked (Shafer and Tversky 1988). The *frequentist* interpretation considers probability to be the number of times a given event will occur in a random sampling of events. With such an interpretation, I might estimate the chances of the bus being late based on the number of past episodes of bus riding when the bus was indeed late. The *propensity* interpretation asks the modeler to consider the causes of a particular outcome. Again, from such a perspective I might estimate the chances that the bus is late based on the fact that it is currently rush hour. Finally, the *betting* interpretation asks the modeler what he might wager on a particular outcome. This last perspective has the benefit of being easy to understand, but presumes that a mathematical-like probability is lurking as a potential somewhere inside the modeler, and that money is the catalyst that will finally allow it to be realized. As Shafer & Tversky point out, this runs counter to empirical observations.

A problem may be decomposed in several ways using Bayesian semantics, but one of the more common and better understood approaches is the Bayesian belief network approach (Pearl 1988). Belief networks represent domain knowledge as a directed graph of variables, which are conventionally assumed to be causally related to one another; that is, each source variable causes its target. Hence, belief networks adopt a propensity interpretation of probability. This interpretation is not required and cannot be applied in every situation; oftentimes, the inclusion of an internal "summary" node which has no phenomenological analog may be necessary to simplify the construction of a model. However, the causal interpretation serves as a useful metaphor for the modeler.

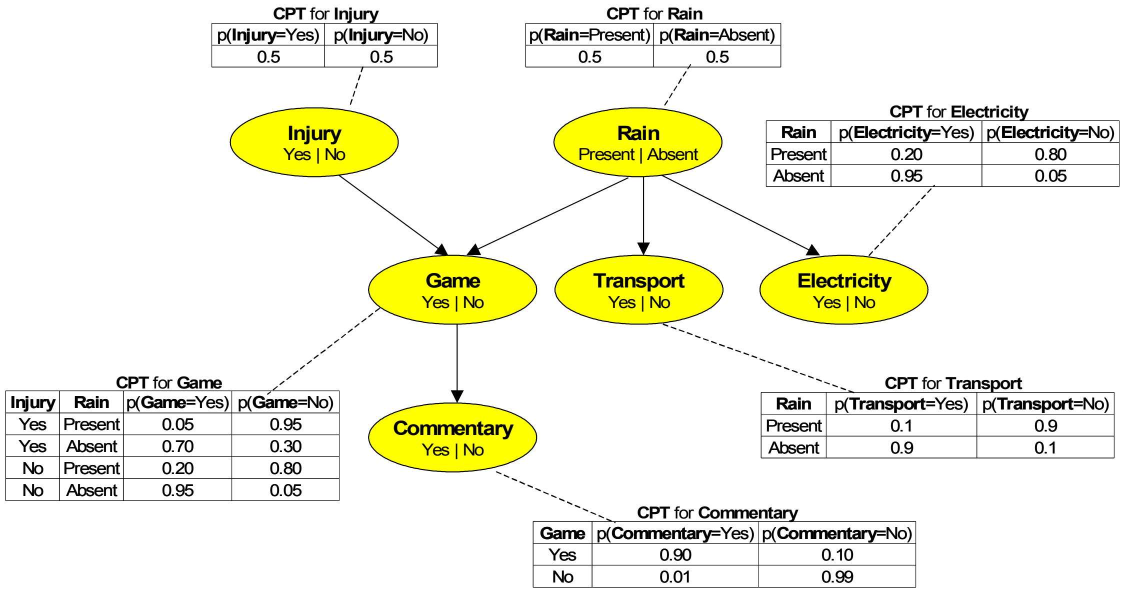


Figure 17: An example of a Belief Network; credit to Dr. Subrata Das

The topology of a belief network captures the modeler's qualitative understanding of the causal relationships between variables. Each variable can have multiple states, which are mutually exclusive and exhaustive. There can be no cycles, and each variable must be an independent source of variance in the model.

An example of a belief network is shown in Figure 17. The belief network serves to estimate the likelihood that the baseball game at Fenway will be played tonight. It encodes the relationships over the domain consisting of the binary variables, *Injury, Rain*, *Game, Transport*, *Electricity*, and *Commentary*; its topology captures the commonsense knowledge that:

- *Rain* causes *Transport* disruption
- *Rain* causes *Electricity* failure
- *Game* causes running *Commentary* on the radio

- *Injury* and *Rain* prevent *Game* from being played

The behavior of the model is quantified through the definition of “prior” probabilities for each variable that has no incoming links, and conditional probabilities relating variables that are connected to one another. The prior probabilities determine the likelihood a variable is any particular state before any evidence is known, and each state in a target variable is conditioned upon the states in the source variables. For any variable that is the target of another, the modeler must provide a conditional probability for each combination of states in its source variables, for each of its own possible states. Thus, for each connected set of variables, the modeler must provide as many probabilities as the product of the number of states in each of those variables.

Probabilities are captured within *conditional probability tables* (CPT) that are associated with each variable in the network. In Figure 17, CPTs for each of the related variables are shown. For example, the probability of having electricity during rain is only 0.2, whereas the probability of having electricity with no rain is 0.95. Note that the CPT for the “Game” variable includes eight entries, which enumerate all possible combinations of “Yes” or “No” and the states of the parent variables.

Once a model is defined, is may be used by “posting” evidence (setting a variable to a given state) based on what is known about the world. Belief updating algorithms (based upon Bayes rule, but generally calculated using optimized algorithms) cause belief to propagate through the model and update the likelihoods of all non-observed variables.

Belief networks are useful and powerful tools for building domain-specific decision support applications. Furthermore, simply extending belief networks by associating utility estimates with states in nodes can turn them into influence diagrams (a close derivative of decision trees), thus completing the SEU model. Unfortunately, constructing belief networks is a labor intensive process that presents even experienced end-users with several hurdles. As is apparent, the number of probabilities that must be defined can be quite large, especially as the number of variables and relationships within the model grow. Within the "interior" of a large network, the semantics of causality can be lost, and assignment of conditional probabilities may become little more than educated guesswork. Iterations with different conditional probability values are often required to achieve the correct kind of behaviors. Furthermore, any modifications to the topology of the network may render many conditional probabilities invalid, and require the elicitation of new values.

I will now turn to Dempster-Shafer theory, another approach to expressing probability that addresses some of the problem with the Bayesian approach.

2.1.2 The Dempster-Shafer Approach

The Dempster-Shafer (D-S) theory of belief functions (Shafer 1976) has not had as widespread an impact on the decision support community as the Bayesian approach. However, as I will describe, it has a number of potential advantages from the modeler's perspective.

As with the Bayesian approach, the D-S theory of belief functions is grounded in the axioms of probability. Unlike Bayesian theory, the unit of analysis in D-S theory is not a probability value, but instead is a *belief function*. A belief function describes the meaning of sets of evidence in terms of the probability any piece of evidence is true. Shafer and

Tversky (1988) describe the semantics of a belief function as:

> Suppose someone chooses a code at random from a list of codes, uses the chosen code to encode a message, and then sends us the results. We know the list of codes and the chance of each code being chosen – say the list is $o_1, \ldots, o_n$, and the chances of o_i being chosen is p_i. We decode the message using each of the codes and we find that this always produces an intelligible message. Let A_i denote the message we get when we decode using o_i. Then we have the ingredients for a belief function: a message that has the chance p_i of meaning A_i (Shafer and Tversky 1988; p. 247-8).

Phrased in terms of "codes" and "messages" D-S belief semantics can seem a little less straightforward than Bayesian probabilities. However, if we consider "codes" as evidence, and "messages" as the meaning of that evidence, mapping the above description into a decision support context becomes somewhat more straightforward. A belief function may be thought of as analogous to a situation in which the modeler has been given the advice of an expert (or testimony from witness); the meaning of that advice is only as reliable as the source from which it originated. Probability indicates the reliability of the source.

A somewhat more formal description may clarify things. Let Ω be a finite set of mutually exclusive and exhaustive propositions, called the *frame-of-discernment*, about some problem domain. $\Pi(\Omega)$ is standard notation for the power set of Ω. A *basic probability assignment* (*bpa*), $m : \Pi(\Omega) \rightarrow [0,1]$, is used to quantize the belief committed to a particular subset *A* of the frame of discernment given some evidence. The probability number $m(A)$ indicates how much belief there is that the correct value is in *A* is in fact the case, where $m(\Omega) = 0$ and $\sum_{A \subseteq \Omega} m(A) = 1$. The value 0 indicates no belief and the value 1 indicates total belief, and any values between these two limits indicate partial beliefs. A basic probability assignment *m* is Bayesian if $m(A) = 0$, for every non-singleton set *A*. For any set $A \subseteq \Omega$ for which $m(A) \neq 0$, *A* is called a *focal element*.

Thus, unlike Bayesian theory, where each state in a random variable must be assigned a unique probability, D-S theory allows belief values to be assigned to sets of states. For instance, it is possible in D-S theory to state the belief "I believe it will either rain or snow" without defining the apportionment of belief between these states.

D-S theory also provides us with a natural representation of uncertainty, by allowing some portion of belief to remain uncommitted. That is, the probability number *p* may be known for only a partial set *A* of hypotheses. In this case, the residual complementary probability number *1-p* is assigned to the frame-of-discernment (Ω) as a whole. This residual value is a representation of ignorance, and can be very useful when modeling decision problems, or adjusting the weight of evidence based on updates to the reliability of a source.

This representation of ignorance allows us to distinguish between *likelihood* and *confidence*. Suppose a coin is flipped four times, yielding 2 heads and 2 tails. Under the traditional Bayesian interpretation of probability the *likelihood* of heads on the next toss is clearly 50%, given the observed data. Now, suppose a coin is flipped 100 times, yielding 50 heads and 50 tails. Again, the *likelihood* of heads on the next toss is 50% - however, we should be far more *confident* in the likelihood computed from the second experiment.

There may be ways to include confidence in the Bayesian perspective (e.g. including an "ignorance" state; this strains the semantics of Bayesian probabilities somewhat), but D-S theory allows us to make this distinction natively. Hence, we might assign a belief of .05 to the likelihood of heads and .05 to the likelihood of tails after the first four flips, indicating that heads and tails are equally likely, but that .9 of our total belief remains unassigned because we lack confidence in the data. After a hundred flips, we might assign .45 to our belief that heads will be next and .45 to our belief that tails will be next, and only .1 of our belief remains unassigned because we are fairly confident in our observations.

Dempster's rule of combination offers a means for combining beliefs about the same frame of discernment from multiple sources. Two independent basic probability assignments m_1 and m_2 can be combined into a single joined basic probability assignment $m_{1,2}$ by:

$$m_{1,2}(A) = \begin{cases} \dfrac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B \cap C = \Phi} m_1(B) m_2(C)} & A \neq \Phi \\ 0 & A = \Phi \end{cases}$$

D-S theory has several advantages over Bayesian theory in developing collaborative decision support applications. Aggregation rules are a natural fit for combining conflicting data from multiple sources, and the confidence in any given belief assignment can be easily modified by proportionally adjusting the belief allocated to focal elements in a basic probability assignment from a given source. Perhaps most importantly from the modeler's perspective, there is no need to elicit large numbers of condition probability values.

There has been relatively little focused work incorporating D-S theory into a modeling framework, like the Bayesian network framework, that can guide its application to decision problems. One of the few approaches is suggested by Das (2005), who describes how the Dempster-Shafer theory of belief functions may be used in combination with an argument formalism. This argument formalism was first introduced as a reasoning scheme for application in an expert medical decision making application (Das et al. 1997), and was designed to be agnostic with respect to the type of belief aggregation method employed.

Table 6 provides an example of this approach. The decision problem is to decide upon a course of action - hold position and wait for reinforcements, attack from the north, or call in air support - in a battlefield situation.

```
Decision:
      Choose a course of action
Alternatives:
        1. {HOLD} Hold position and wait for reinforcements
        2. {ATTACK} Attack from the north
        3. {AIR SUPPORT} Call in air support
Arguments:
      Reinforcements are days away ➔ support (not HOLD, .5)
      Fewer casualties ➔ support (HOLD, .5)
      Heavy casualties ➔ support (not ATTACK, .7)
      Surprise ➔ support (ATTACK, .5)
      Resource usage ➔ support (not AIR SUPPORT, .4)
      Civilian casualties ➔ support (not AIR SUPPORT, .7)
Commits:
      netsupport(X,M) & netsupport(Y,N) & netsupport(Z,O) & M >
      N & M > O ➔ decide (X)
```

Table 6: A decision example

Arguments are written according to an *argument schema*. In general an argument schema is like an ordinary inference rule with

```
support(<candidate>, <sign>)
```

as its consequent, where `<sign>` is drawn from a set called a *dictionary*. The `<sign>` represents, loosely, the confidence that the inference confers on the candidate. The dictionary may be strictly quantitative (e.g. the numbers in the [0,1] interval) or qualitative (e.g. the symbols {+, -} or {pro, con}). In the above example, we are dealing with probabilistic arguments and `<sign>` is drawn from the probability dictionary [0,1]. Thus, in Table 6 the argument,

```
Civilian casualties ➔ support (not AIR SUPPORT, .7)
```

means that there is a 70% chance of civilian casualties if air support is called in, and this argues against that alternative.

The "Commits" portion of the argument describes how a decision is to be evaluated. In the example shown in Table 6, a single commitment rule selecting the alternative with the highest support will be chosen. The aggregation method to be employed is represented by the netsupport function, which aggregates collections of arguments for and against any candidate to yield a measure of the overall strength for it.

It is a simple matter to apply D-S theory to the above model (by plugging Dempster's aggregation rule into the netsupport function), and Das (2005) briefly describes this approach. Note, however, that as stated, the model only allows for a single level of analysis – that is, each argument may be related to an alternative, but there is no opportunity to argue about an argument. In the following chapter I will describe a simple extension that will support this kind of analysis.

2.2 Utility Models

The other part of the SEU model, and another apparent "irrationality" in human decision making, is the expression of a consistent set of preferences. Utility models are designed to help people overcome these difficulties. The majority of utility models may be classified under the rubric of Multi-Criteria Decision Analysis (MCDA). All MCDA approaches: 1) break a decision problem down into a set of non-overlapping alternatives; 2) identify a complete a set of (possibly hierarchically organized) criteria to be evaluated for each of the alternatives; 3) assign weights to each of the criteria, and; 4) evaluate these criteria to establish a ranking over the alternatives. Differences across MCDA approaches include how hierarchies (if they exist) are represented, how criteria are established, and how the results are ultimately aggregated to establish a ranking. There are many such models; Buede and Maxwell (1995) offer an excellent starting point for the interested researcher. I review one MCDA model here, called the Analytical Hierarchy Process (AHP) to provide the reader with a flavor.

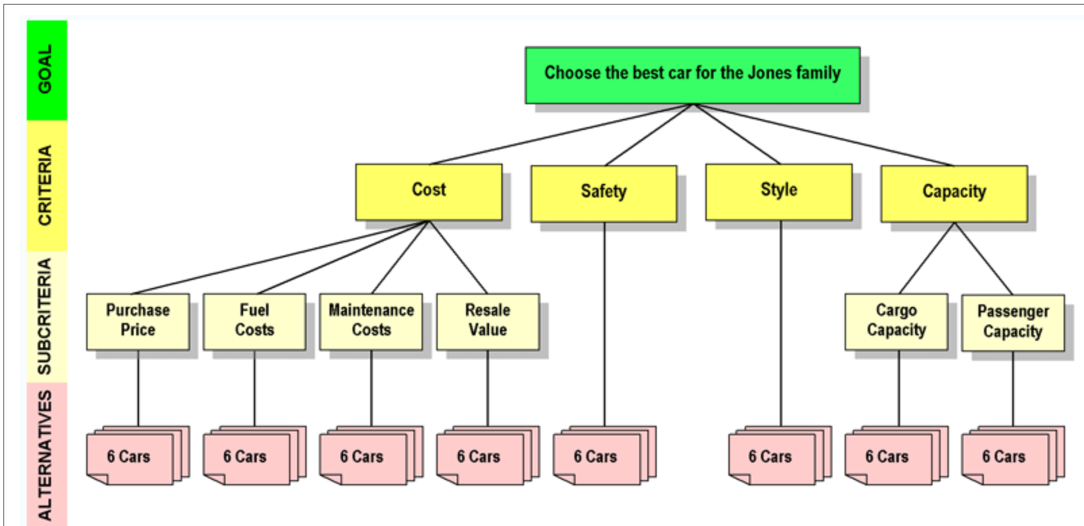


Figure 18: Problem decomposition in the AHP; credit to Lou Sander.

The AHP (Saaty 1990) is unique in that is establishes a ranking for the criteria by eliciting pairwise comparisons among all of the criteria at each level in the hierarchy. As with other MCDA models, a problem is decomposed into a set of alternatives, and a set of criteria along which the alternatives may be compared. Figure 18 illustrates a sample decision problem, describing the factors involved in choosing between six cars. The "goal" is a statement of the decision problem ("Which car should we buy?"), and the criteria are the various factors involved (e.g. cost, safety, etc.). As illustrated, criteria may be decomposed into sub-criteria (and further, if need be), but only the leaves in the criteria hierarchy will be evaluated with respect to the alternatives (labelled "6 cars" in the figure). During the aggregation process, and after criteria ranking have been established at each level in the hierarchy, those criteria in the interior of the hierarchy are collapsed into their children by taking the product of any leaf criteria and all of its ancestors.

$$\begin{array}{c c}
 & \begin{array}{cccc} \mathit{cost} & \mathit{safety} & \mathit{style} & \mathit{capacity} \end{array} \\
\begin{array}{c} \mathit{cost} \\ \mathit{safety} \\ \mathit{style} \\ \mathit{capacity} \end{array} & \begin{bmatrix} 1 & 1/3 & 5 & 3 \\ 3 & 1 & 7 & 5 \\ 1/5 & 1/7 & 1 & 1/3 \\ 1/3 & 1/5 & 3 & 1 \end{bmatrix}
\end{array}$$

Figure 19: Relative rankings for criteria in the top leve of the hierarchy

To determine values for each of the criteria (and this is the unique and critical step in the AHP) the modeler ranks the relative importance of each pair of criteria in each level, which produces a reciprocal matrix (symmetrical entries are reciprocal). An example of a ranking matrix for the first level of criteria shown in Figure 18 is given in Figure 19. Each entry in the matrix may be interpreted as "the degree to which each <*row heading*> is more important than <*column heading*>." Thus, "safety" is three times as important as "cost," but "cost" is five times as important as style. Computing the overall priority rankings for each of the criteria is accomplished by finding the eigenvector of the matrix. Saaty also provides a technique for determining the degree of inconsistency (e.g. intransitivity) in the elicited criteria rankings. Too much inconsistency serves as an indicator to the analyst that criteria rankings should be re-examined. This procedure is applied from top to bottom of the hierarchy to arrive at the "best" alternative.

Many other utility estimation methods have been proposed and evaluated, but they generally follow a similar pattern – preferences are identified, occasionally in a hierarchical fashion, smaller judgments are made about subsets of criteria, and an algorithm is used to create global preferences. When applied in the context of actual decision making, these processes are advocated as tools that encourage careful thought about a problem. The net effect, however, is that they help people to arrive at a consistent arrangement of preferences that is compatible with the SEU paradigm.

In the preceding sections, I have described some of the more common approaches to representing and modeling decision problems. These models suggest an "ideal" with respect to how information should be combined to arrive at a decision, and it precisely these ideal which appears to be violated in the common knowledge problem. That is, people do not appear to combine all of their available information to arrive at an outcome that would be predicted by these models. Thus, the rational model will serve as the basis for the adaptive mediation I would like to include in my envisioned platform.

I will now turn my attention some of the tools that have been built based on the rational model, and discuss some of the reasons why these tools have not met with much success, and an approach to remedying the problem.

3 Group Decision Support Systems

Tools to support group decision making draw together research on collaboration and decision analysis. Groupware features are incorporated to overcome common process losses (like production blocking and airtime fragmentation), and decision analytical techniques are added to help people structure their decision making process. In their early taxonomy of group decision support systems (GDSS), DeSanctis and Gallupe (1987) divided GDSSs into three levels:

1. Level 1 GDSSs – Provide technical features to overcome communication problems. Such platforms may include voting, but little else in the way of specific decision support technology. As such, the primary benefit of a Level 1 GDSS is its ability to counter process losses in group discussion.

2. Level 2 GDSSs – Equip Level 1 GDSSs with decision analytic tools for keeping track of, organizing, and evaluating decision relevant information.

3. Level 3 GDSSs – Add further, machine driven process structure to Level 2 GDSSs. For example, a Level 3 GDSS might impose parliamentary procedures upon a group discussion, or automatically prompt users for input based upon some inference process.

At the time DeSanctis & Gallupe proposed their taxonomy, there were few Level 3 GDSSs in existence, and this remains the case today. Such systems have probably not been widely explored because early experiences with systems like the Coordinator (Flores et al. 1988) demonstrated that people did not respond well to collaborative work that was too rigorously structured.

In the following sections, I will focus upon Level 2 GDSS systems. These systems are a kind of adaptive groupware that, in my taxonomy of adaptive systems, offer domain-oriented goal adaptation. That is, these systems are equipped with an algorithmic model that embodies an idealized mathematical description (to varying degrees) of a "sound" decision, and the users add their information to the system in order to attain a product, which is an indicator of the "best" decision.

3.1 Level 2 GDSS

The majority of Level 2 GDSSs incorporate decision support features that are based in the rationalistic tradition of decision analysis. These systems typically embody a particular mathematical modeling approach that is consistent with SEU decision theory. They guide collaborators in identifying the relevant components of the model, and assigning quantitative metrics to each of these components.

Likelihood models are not generally a part of these systems, and the few systems that have attempted to incorporate them have either not fully materialized, or are not suitable for general purpose group decision support. For example, the CoRaven (Jones et al. 1998) platform integrated belief networks for intelligence analysis in a collaborative problem solving platform. Different members of a collaborative team would select and modify portions of belief networks for intelligence analysis. "Structure advisors" on the team would select and / or modify portions of belief network structure, and "likelihood advisors" would assign probabilities to this network. However, there is no published work suggesting that the platform was ever fully implemented or deployed.

Deshpande, de Vries, and van Leeuwen (2005) also describes an experimental system that supports collaborative belief network construction for decision support. Belief networks are employed as a mediating artifact designed to clarify and support the construction of common ground in problem solving. Users interact by first creating iconic representations of data relevant to the problem in a "collaborative problem space," then specifying causal relations, and finally creating conditional probability tables for each node. The platform offers a novel approach to constructing belief networks, but because it assumes a fair amount of expertise with belief networks on the part of the collaborators, it unlikely to be of much use for general purpose decision making.

By and large, most GDSS tools adopt an MCDA approach, which tends to turn a blind eye to the issue of uncertainty. These platforms vary according to the degrees to which they emphasize support for the collaborative process versus problem modeling. GroupSystems' ThinkTank lies at the collaborative process end of the spectrum. ThinkTank (version 2.0) is the most recent instantiation of a system that originated with the GroupSystems platform originally developed at the University of Arizona (Nunamaker et al. 1991). ThinkTank scaffolds the group decision making process through the following series of steps:

1. Brainstorming – Participants work simultaneously to generate ideas, proposals, or

alternatives. ThinkTank provides a chat-like tool that allows participants to discuss each of the ideas.

2. Organization – Participants work simultaneously to "bin" ideas into categories.
3. Prioritization – For each idea / alternative, any number of criteria may be specified. For instance, each alternative can be associated with an "importance" and "ease of implementation" criteria. Collaborators are able to vote on each of these. ThinkTank then provides several tools for analyzing these outcomes; the tool indicates which alternatives have the highest value as well as the amount of consensus for each of these.
4. Planning / Report Generation – The system provides tools for gathering external feedback from stakeholders (via automatically constructed surveys) and summarizing results as a plan for action.

The underlying mathematical model for utility estimation is rudimentary – it is an ad hoc MCDA tool, with an organizational hierarchy imposed on top of the selection of alternatives. There is little explicit guidance within the system as to how "alternatives" and criteria should be selected, and its analytical tools offer only simple techniques for examining the distribution of votes over alternatives and criteria (mean and standard deviation). The emphasis in ThinkTank is upon its rich collaborative features, which are designed to allow many people to work in parallel, and overcome the process losses described above. The analytical tools serve as a resource for further collaboration.

At the other end of the spectrum are packages like the Decision Lens (http://decisionlens.com; accessed 07/08) and Expert Choice (http://www.expertchoice.com; accessed 07/08), both of which employ the AHP as their underlying methodology. Group decisions in AHP can be achieved by soliciting individual criteria assessments from all users, and then combining them via a geometric mean (Saaty 2001; pg. 61). In one suggested mode of operation, individuals use wireless keypads to submit their rankings directly to the system. In this manner, the group functions as a disjoint set of inputs to the model. Once the model has been established, no further negotiation between collaborators is required.

Such tools differ from tools like ThinkTank in that the emphasis is on the mathematical model, rather than the collaborative process. In either case, some interaction will be necessary to establish the set of alternatives and criteria to be evaluated. In the case of ThinkTank, this process happens initially in parallel, but subsequent discussion is supported though the system via message board discussions that may be attached to every "idea" generated. Analogous support is not available in DecisionLens, and it is strongly recommended that a facilitator be employed during the model construction phase.

Once a model is generated in either platform, voting and analysis may indicate a need for further refinement. In ThinkTank, it is envisioned that this will entail further iterative cycles of collaboration. In DecisionLens, this may either be a matter of adjusting individual rankings to eliminate inconsistencies, or revising the model. If revisions to the model are required, an external facilitator may be necessary. Once again, the emphasis in the DecisionLens is upon limiting interaction to that which is necessary to achieve a sound mathematical model.

Despite sustained interest and years of work, Level 2 GDSS remains something of a disappointment. Hundreds of empirical studies have been performed with GDSSs of varying types (see Fjermestad and Hiltz 1998 and Fjermestad 2004 for comprehensive reviews). These have shown that GDSS have the potential to substantially improve group processes (especially with respect to process losses), but this potential is realized far less frequently than might be hoped (Briggs 2006). Level 2 systems have not been broadly adopted, either. Limayem and DeSanctis (2000) suggest that "a key barrier to the use of MCDM ... is their sheer complexity" (pg. 386).

These issues reflect a dimension of the knowledge acquisition problem. On the one hand, systems like ThinkTank attempt to relax constraints on the underlying mathematical model and support active collaboration, but do so at the cost of high-quality data that might be used to drive a powerful analytical engine. This lack of an analytical result might lead some to question the platform's value.

On the other hand, systems like DecisionLens carefully control collaboration in order to get the kind of information that is required to produce a "rational" product. In this case, the effort required to get necessary information into the system may present too great a hurdle for the users.

This double bind is exactly the sort of hurdle faced by developers of adaptive systems. To make matters even more difficult, the payoff for using such systems, whether or not they produce an analytically sound result, is in question. There are unfortunately no studies of the form "twenty companies using platform X made significantly more correct decisions than twenty companies using platform Y." Furthermore, and as exemplified by the discussion of the hidden profile studies in Level 1 GDSS systems above, empirical research with GDSS systems has been unable to identify a single, reliable, consistently reproducible improvement in small groups using GDSS systems (see Briggs 2006).

A questionable payoff, coupled with an unnatural and laborious interaction, is part of the reason why GDSS systems have not made much headway in the world. To address these problems, some researchers have proposed the addition of automated facilitation to guide people in the use of the technology (see Lagroue 2006). Based on my analysis, and the work described in previous chapters, a more direct solution may exist if we can address the common knowledge problem.

To do this, we would like to introduce a CR that helps people better coordinate their deliberative process, and then use the information made available by such a CR to adapt collaborative decision making towards the rational ideal described above. To find support for such a CR, I turn to a body of research that is specifically concerned with supporting deliberative dialog.

4 Argumentation

Argumentation is a very broad topic of investigation, primarily concerned with how conclusions may be reached from a set of premises through the application of reason. Scholarly research may be found that connects argumentation to most academic fields of investigation, including philosophy (Toulmin 1958), psychology (e.g. Voss, Tyler, and Yengo 1983), law (e.g. Newman and Marshall 1991), education (e.g. Kuhn 1991; Carr 2003), and AI (e.g. Dung 1995; Das et al. 1997). In decision making contexts, argumentation is designed to support the reasoning process rather than the judgment process.

Argument theories may offer a formalism and a set of rules designed to guide practitioners to the creation of a well reasoned argument. Argument formalisms may or may not be grounded in a mathematical theory, but the formalisms themselves do not provide a quantitative means for dealing with likelihood or utility. Thus, theories of argument do not fit within the rubric of SEU approaches to decision analysis. However, an argument formalism can serve as a guide to a “well-reasoned” organization of information in support of (or in objection to) a choice among options, and so can be considered a prescriptive model of decision making. Whether or not it might also be considered a normative model is perhaps a matter of argument.

In recent years, a number of tools based on argument formalisms have been developed as a means for facilitating group decision making. These systems are designed help make deliberation more effective. As they are not approaches within the rationalist tradition, they make no attempt to specify utility or likelihood, nor do they “aggregate” exchanged information to arrive at a best answer. Such systems may be more accurately considered structured dialogue mapping tools, because they introduce semi-formal representations that are used to build graphical representations of the reasoning process that underlies deliberative dialog. These platforms are collectively referred to as computer-supported argument visualization (CSAV) platforms (Buckingham Shum 2003).

CSAV systems typically claim lineage from either IBIS (Kunz and Rittel 1970), or Toulmin's theory of argument (Toulmin 1958). They offer varying support for mediated communication, but are usually considered to be collaborative tools. Where support for mediated communication is not included in the platform, is it envisioned that deliberation will take place in a face-to-face setting.

Unlike MCDA platforms, the bulk of research that has occurred with CSAV platforms have been within educational contexts, and this has lent a different flavor to the kind of empirical data that exists. Specifically, this literature focuses upon knowledge transfer, integration, and retention in groups. As this body of research grows, proponents of CSAV are beginning to be able to make stronger cases for its utility in "real-world" decision making.

In the following review, I cover the two basic classes – IBIS and Toulmin based - of CSAV systems. In each case, I provide a brief description of the formalism itself, and relevant empirical data where it exists.

4.1 IBIS

IBIS (Kunz and Rittel 1970; Rittel and Webber 1973) is a methodology for structuring issue-based dialog. IBIS was initially proposed as a hypertext environment designed to be used to grapple with "wicked" problems, and specifically wicked problems associated with establishing design rationale. Wicked problems are those problems which include features such as vague goals, constraints that are hard to identify, multiple plausible solutions, and no stopping criteria (Rittel 1972).

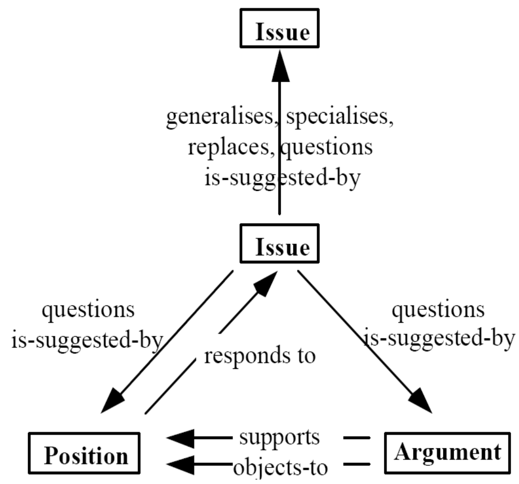


Figure 20: Basic IBIS structural units and links

As a representational scheme (shown in Figure 20) IBIS introduces three statement types (Issues, Arguments, and Positions) and eight relationship types (responds-to, questions, supports, objects-to, specializes, generalizes, is-suggested-by, and replaces). Issues have the form of questions, and can be further sub-classified as:

— Factual issues – "Is X the case?"

— Deontic issues – "Shall X become the case?"

— Explanatory issues – "Is X the reason for Y?"

— Instrumental issues – "Is X the appropriate means to accomplish Y in this situation?"

It was envisioned that an implemented system would also have links to factual information; for instance, a "factual issue" could refer to a document or expert testimony. Additionally, issues, topics, evidence, etc. would all be organized in IBIS subsystems so that it would be easy to search and reuse information that was compiled.

IBIS was designed to be an implemented formalism. As a result, there is very little research that explores the impact of IBIS as a “pure” formalism, and most research with IBIS has been coupled to the examination the platforms that implement it.

gIBIS was the first collaborative platform to use IBIS as a semi-formal interface representation. gIBIS was technically advanced for its time, and supported true concurrent control and links to external resources. Conklin and Begeman (1988) provided one of the more complete analyses of gIBIS in context. The system was used as an organizational knowledge management tool, rather than as a tool for isolated collaborative decision making episodes. Although their were some limitations with the formalism itself, the platform was very well received, both as a collaborative tool and as a tool for organizing individual thought.

gIBIS evolved into the commercial Questmap platform, which has since been reincarnated as the open-source Compendium™ platform. Compendium™ has added more advanced technical features, including transclusive linking (linking between maps) and support for fully distributed libraries of argument maps. Furthermore, while the default formalism remains grounded in the IBIS methodology, it is fully customizable, so that the user might create whatever special purpose formalism is required.

Unfortunately, there have not been any recent case-studies that describe the impacts of the IBIS formalism on collaborative information processing, perhaps because it is primarily studied and used as an applied technology where research is difficult to perform. Conklin (2003) offers some anecdotal reports on an extended study using the Questmap platform. He observes that users generally find the platform useful once they become proficient with the system, but that there is a long learning curve, and an organization will probably require a management level “cheerleader” to advocate for use of the platform. Buckingham Shum et al. (2006) examine these issues in depth, and arrive at a conclusion that while IBIS based tools will necessarily entail some learning on the part of their users, there is sufficient evidence that the payoff is well worth the investment. A strategy moving forward would be to invest in techniques to facilitate novice use in order to “get them up to speed.”

Presently, though, IBIS is not a tool that is well suited to solving singular decision problems quickly, and it is hard to see how IBIS might fit the criteria for a CR embedded in a GDSS platform. Toulmin's formalism, however, is somewhat more promising.

4.2 Toulmin

One well-known argument formalism comes from Toulmin's theory of argument (Toulmin, 1958). Toulmin developed his theory to address observations about how difficult it is to cast everyday practical arguments into classical deductive form. He claimed that arguments needed to be analyzed using a richer format than the simple if-then form of classical logic. In Toulmin's model, an argument has the form of a logical rule, but not its force (i.e. an argument does not sanction a definite conclusions.)

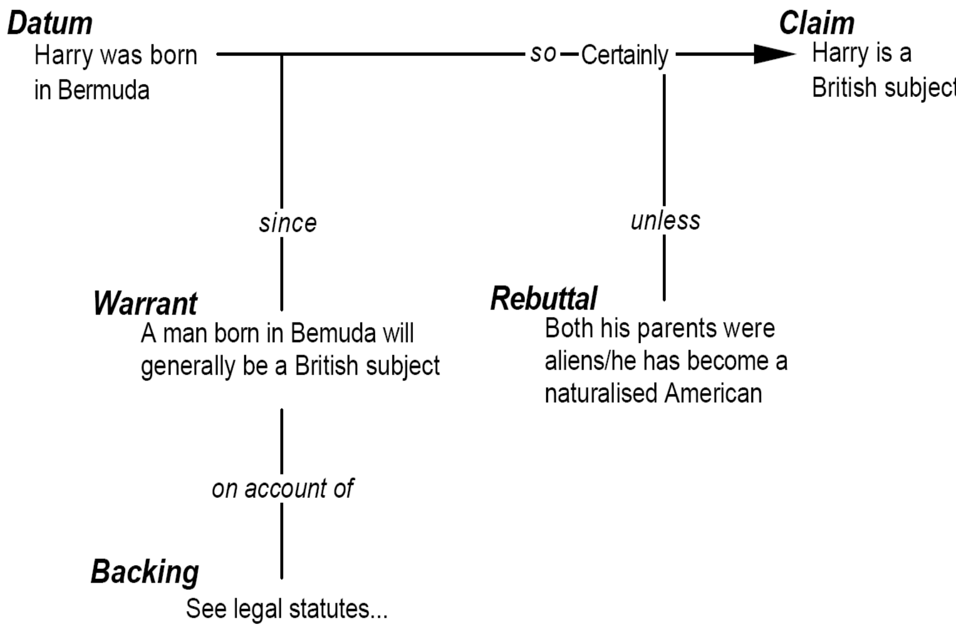


Figure 21: Toulmin's argument formalism; from Toulmin [1956]

Toulmin's model decomposes an argument into a number of constituent elements, shown in Figure 21. The *datum* represent the facts that support the arguer's position, and the *claim* the position that is supported. The *warrant* is the inference rule that allows the datum to support the claim. It does not necessarily have to be a simple causal rule, nor do its predicates necessarily need to match the datum and claim. These three elements form the core of Toulmin's model. In Toulmin's original formulation, the remaining elements were recognized as secondary, possibly implicit, elements that modify the above relationship. The *backing* is the body of knowledge that supports the warrant. For example, in legal contexts, backing may be the body of case law which constitutes "precedence." The *qualifier* is somewhat ambiguous, but provides a taxonomic means for modifying the force of the argument. Finally, the *rebuttal* acknowledges exceptions or limitations to the argument. It was not (initially) intended to be a path along which counter-arguments would be developed, but was rather a means by which the arguer could forestall objections.

Toulmin’s model has been extended in a variety of ways, and used to examine reasoning in several contexts. For instance, Voss (Voss, Tyler, and Yengo 1983; see also Voss 2005) extended Toumin's model to support chaining of arguments (for instance, allowing the claim of one argument to become the warrant of another), and used the extended model to analyze speak-aloud protocols from experts solving “wicked” – or in Voss's terminology, ill-structured – problems. In his analysis, problems were chosen from the social sciences domain, such as “Propose a strategy for increasing agricultural productivity in the Soviet Union.” Voss, Tyler, and Yengo (1983) used transcripts that were coded with Toulmin's scheme to illustrate differences between expert and novice problem solving. Among the differences noted, it was found that experts with training in the social sciences (though not necessarily with a background in the specific problem) devoted more time to developing a problem representation, identified an abstract cause, and proposed few solutions. Novices, and experts in non-social-science fields (e.g. chemistry) did not devote much time to problem representation, identified many concrete causes and connected each cause to a discrete solution. Voss also noted that while the formalism captured the micro-structure of the developed arguments, it did not capture differences in the high-level reasoning strategies. This may imply that if the formalism were applied as a prescriptive technique, it might be used to make the reasoning process more explicit, but that this would not necessarily guide someone to a more “expert” solution.

Newman and Marshall (1991) offer a similar analysis of legal argument, and their results confirm and extend Voss' observations about the limitations of Toulmin's model. As in Voss' work, Newman & Marshall extended Toulmin's scheme to support various means of chaining argument forms, and used the extended formalism to encode arguments made in a Supreme Court case about warrantless search in mobile homes (People vs. Carney). In deriving their extended argument scheme, they clarify some of the difficulties with the function of the rebuttal. Rebuttals, they note, can be applied to the datum, backing, or warrant itself. Furthermore, they observe that rebuttals are "holes" in the representation, and can depend upon contextual information that is not directly part of the represented argument itself - for instance, when the internal logic of the argument is sound, but the conclusion not practically applicable.

This observation becomes part of a larger critique about the inability of Toulmin's argument forms to represent "macro-structures" which guide the reasoning process. Their critique is analogous to Voss's brief observation that high-level strategies could not be captured in the formalism. Newman & Marshall note that backing may be left unspecified because it is part of the common knowledge of the interlocutors. They also observe that while analogies (which serve an especially important role in reasoning from precedent in legal domains) may be forced into Toulmin's form, the resultant representation obscures important structural similarities between the analogues. Finally, argumentation is a goal-directed activity, and must be evaluated against a background of priorities and strategies that cannot be captured in Toulmin's representation.

Newman & Marshall offer some suggestions for "second-order" representational extensions to help capture the macro-structure of the argument. One is a matrix, comparing alternative cases across criteria, to help evaluate a particular argument. Another is a hierarchical list of major issues that should be considered in constructing a particular argument.

Toulmin's theory of argument has had an enormous impact on many disciplines. However, Toulmin's formalism, as it was initially described, is not easy for novices to understand (primarily because warrants are generally left implicit in everyday reasoning). However, as tools have been built around the formalism, it has been simplified to a point where it is often a highly usable structure.

Generally, Toulmin-based platforms frequently simplify the theory to eliminate the need for explicit warrants, and extend the theory to support argument chaining. For example, the Reason!Able™ platform (van Gelder 2003; currently available commercially as Rationale™ from http://austhink.com) consists primarily of "Positions" (claims), "Arguments" which can be "Reasons" for (datum or warrant) or and "Objections" to (rebuttals). The current version of Rationale also contains a separate "rebuttal" element which may be attached to another argument, instead of a claim. Note that while the "warrant" relationship is not made explicit, something close can be achieved by chaining arguments for or against in a series of implications. The role played by the "backing" element can be accomplished by adding an explanation to a leaf node in a special evaluation mode.

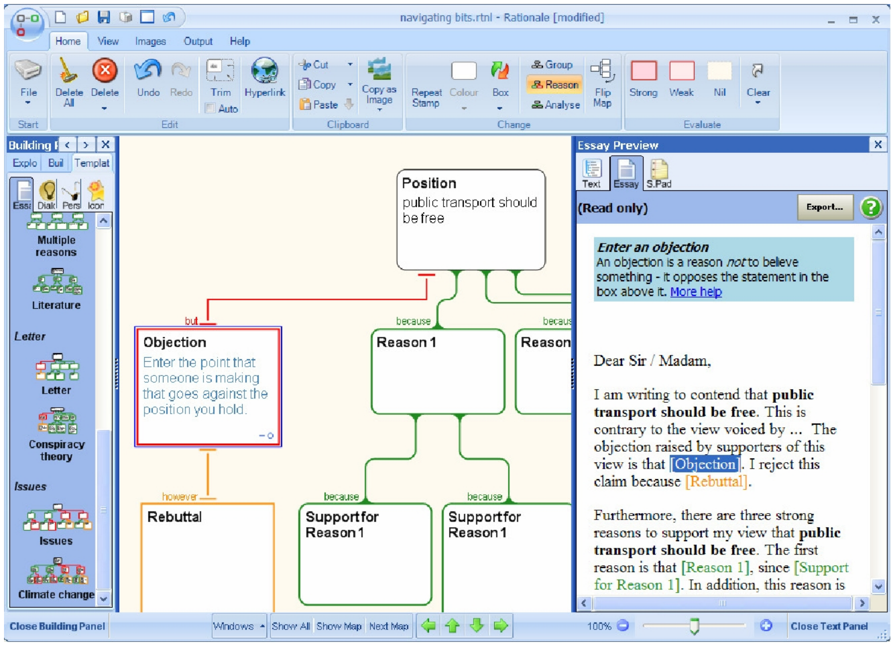


Figure 22: The Rationale Interface

While there is no algorithmic means for aggregating belief throughout a constructed argument network in Reason!Able™, weights may be manually assigned to each argument node (on a four point scale for each pro and con), and arguments may be grounded using one of eight possible labels (such as “Common knowledge,” “Expert opinion,” etc.).

Reason!Able™ was not designed to mediate collaborative discussion directly, nor are its successors. It has been used as an argumentation aid in educational settings, and to settle corporate disputes in face-to-face, facilitated meetings. There is some quantitative evidence that, as a learning aid, Reason!Able™ is able to enhance students' critical thinking abilities (Twardy 2004; van Gelder 2003). Van Gelder (2003) also provides some anecdotal evidence regarding the utility of the software for corporate dispute resolution.

Another platform that is loosely based on Toulmin's formalism is the Belvedere platform (Suthers and Weiner 1995; Suthers 1999; Suthers 2003). Unlike Reason!Able, Belvedere is designed to support mediated discussion. Belvedere was initially based on a complete instantiation of Toulmin's theory, but students exhibited confusion about the meaning of the provided primitives. This confusion interfered with the the students' ability to communicate (Suthers 1999). These difficulties led Suthers and his colleagues to simplify the set of primitives to choose from. The reduced palette contained “data” and “hypothesis” node types, and “consistent” and “inconsistent” link types. Subsequent investigations have not examined the usability of this reduced formalism directly, but empirical studies with this (and later) versions of system have found that the representation facilitates consideration of evidential relations (Toth, Suthers, and Lesgold 2002). Other studies have found that the graphical interface presentation (as opposed to a text based interface) plays a significant role in mediating communications in distributed settings (Suthers, Girardeau, and Hundhausen 2003).

One empirical study with Belvedere examined the effect of different interface representations in a hidden profile decision task (Suthers et al. 2008). Three versions of the interface were compared. One version provided users with a threaded chat interface, which did not employ an argument formalism. Another version used Belvedere's graph-based argumentation interface, with primitives for “notes” and “unspecified” nodes, in addition to those described above. A third version had both threaded chat and graph interfaces, and elements in the graph could be linked to locations in chat.

Participants in all conditions shared roughly the same amount of information, but participants in the graph condition had significantly more consensus in the final judgment (as measured by the convergence of their individual judgments in post-tests). One interesting result was that groups using the graph condition produced more hypotheses, and both graph and mixed conditions elaborated their hypotheses more than the text condition. Groups in the graph condition were also better able to integrate and draw conclusions from this information in post-tests than the other groups. Suthers used these results to support his hypothesis that a representation that makes conceptual objects and relations explicit would improve collaborative knowledge construction.

Some aspects of the experimental design used by Suthers et al. (2008) limit the relevance of the results to more traditional hidden profile investigations, which are designed to investigate knowledge pooling more directly. For instance, it examined only dyads in a simulated asynchronous environment. Only one piece of 48 information pieces was shared by the participants, and this information was fed to participants gradually over the course of the study. It is not clear if the participants were ever made aware of the valid possible hypotheses, or when they made aware of them. However, the results do provide a clear indicator that the argument representation had a generally positive impact group information processing.

Some researchers have explored the use of Toulmin's formalism as a tool for teaching rhetorical skills. However, empirical studies of the effectiveness of such training have been inconclusive. Carr (2003) describes an experiment with an online collaborative argumentation tool called Questmap that allowed the construction of arguments that adhered to Toulmin's formalism. In the study, Carr compared the performance of second and third year law school students that used Questmap as a study aid to those that did not. Carr found no significant differences between the two groups. One explanation offered by Carr, echoing similar observations by Suthers (1999) is that in practice, the tool did not mediate the collaborative argument construction process. Rather, participants would use the tool as a transcription aid instead of collaborating "through" the tool. It is unclear if the tool would have had more of an impact if it had been a stronger mediator.

Toulmin based systems seem to have the right kind of properties we look for in a CR. That is, they help people to coordinate their deliberative process, and users like to use them. Furthermore, the formalism they employ is very similar to the one introduced by Das (2005) as an argumentation framework that is compatible with Dempster-Shafer belief aggregation.

A couple of systems have sought to combine rational decision models with argumentation style interfaces, which is exactly what I seek to do with mine. I will discuss those in the following section.

5 Hybrid Systems

While there are not many, a couple of systems have married argumentation-like interface representations with SEU style belief aggregation. An early example of this was the SIBYL platform (Lee 1990). SIBYL introduced an argument formalism called DRL (discourse representation language), which was very similar to IBIS. DRL included representations of goals and sub-goals, alternatives, claims, "supports" relationships, "denies" relationships, and other primitives. At its inception, SIBYL was intended to be compatible with a variety of algorithms for handling uncertainty and computing weights, including Bayes and Dempster-Shafer theories. It is not clear, however, that such support ever materialized.

More recently, SRI's SEAS platform is built around the notion of structuring knowledge elicitation and capture using an argument formalism (see Lowrance et al. 2008 for a recent description). The formalism itself is not based upon any traditional approach to argument – rather, it is a hierarchy of topically organized questions, arranged much as in a standard MCDA decision tool (see Figure 23 for an example of a hierarchy). Each interior node in the hierarchy specifies a fusion rule, which determines how its children will be aggregated to arrive at a value. At the base level of this hierarchy (and only at the base level), each question is associated with a multiple choice answer, to be entered by an individual with the relevant expertise. The system automatically aggregates values associated with these answers through the hierarchy according to the fusion rules to arrive at a top level answer.

1. **POLITICAL:** Is this country headed for a political crisis?

1.1. **POLITICAL INSTABILITY:** Is political instability increasing?

1.1.1 **INCREASINGLY UNSTABLE/WEAK GOVERNMENT:** Is the government becoming increasingly unstable or weak?

1.1.2 **INCREASING CONFLICT OVER POLICY/ISSUE AREA:** Is increasing conflict over policy/issue areas having a destabilizing effect?

1.1.3 **DECREASING PUBLIC CONFIDENCE:** Is decreasing public confidence in the leadership or government policies having a destabilizing effect?

1.2. **POWER STRUGGLE:** Is there a government power struggle with potentially destabilizing consequences?

1.2.1. **FACTIONALISM:** Is there evidence of growing factionalism within the government, bureaucracy, or legislature that is leading to or exacerbating a power struggle?

1.2.2. **OPPOSITION CHALLENGE:** Is there a significant political opposition challenge to the government that is leading to or exacerbating a power struggle?

Figure 23: Example hierarchy of questions in SEAS

The focus of SEAS is upon establishing and reusing knowledge within an organization, rather than traditional group decision support. It is designed to be accommodating of different roles within an organization. Every object in every argument is assigned a unique identifier, and access rights can be established for each of these. Arguments are private until published, but arguments may be co-authored; collaboration among authors is accomplished via explicit locking and merge operations. Within ongoing analyses, authors may post general notes or flags to indicate that attention is required on a given item. Argument templates may be stored with links to external information resources that can be used by others to help answer the questions posed by the template.

While SEAS is illustrative of a way in which argumentation and more analytical models may be combined, it does not offer the kind of support for deliberative engagement offered by some of the Toulmin-based formalisms. Hence while the SEAS interface may serve as a CR designed to solve the problem of organizational knowledge management, it does not support collaborative discussion.

In the following section, I will summarize all the material that been covered, and describe the skeleton of a group decision support system that can address the common knowledge problem.

6 Summary

The goal of this chapter was to lay the groundwork for a system that would leverage information made available by a CR, and use this information to add adaptive mediation to a groupware platform in order to eliminate dysfunction in the collaborative process. To this end, I have answered the three questions proposed at the outset of this Chapter, as follows:

1. **Question**: What is the problem with the collaborative process that we wish to address?

   Answer: The common knowledge problem. Specifically, the fact that even when collaborative decision makers have access to information that would lead them to the "correct" decision, they ignore information that they do not share.
2. **Question**: For the problem we've identified, what is the ideal collaborative process?

   Answer: The rationalistic model describes what the collaborative outcome should look like. A variety of techniques are available.
3. **Question**: What type of CR can we use to solve the knowledge acquisition problem?

Answer: Toulmin-based argument visualization platforms have been very well received, and appear to improve people's ability to coordinate their deliberative dialog. Such formalisms also appear to be compatible with Dempster-Shafer theory.

The answers to these questions provide a framework for the development of a group decision support platform that can adapt the group decision making process to make it more “rational.” The envisioned system will employ an argument visualization interface, similar to Belvedere and Reason!Able. The collaborators' use of the interface will be used to drive an engine that is based on Das' (2005) suggested approach to using Dempster-Shafer theory to evaluate argument networks (although some extensions will be necessary). All that is left to determine is how the pieces will be put together to adapt the group process. That is described in the following chapter.

Chapter 4: The Design of REASON

The previous chapter set the stage for a mediated group decision support platform that can address the "common knowledge" problem. The common knowledge problem describes two related phenomena in group decision making. The first is the tendency of groups to exchange more shared information (information that all members have prior to discussion) than unshared information (information uniquely held by some individuals); the second is that groups seem to make decisions that are better predicted by the distribution of individual opinions than by the information that they exchange.

Empirical evidence has shown that there are at least two parts to the common knowledge problem – information exchange that is dominated by shared information, and group information processing that does not seem to be based on exchanged information. Prior empirical work has shown that GDSSs can be used to ameliorate the former problem. GDSSs appear to increase overall communication bandwidth to the point where more unshared information is exchanged than would be in face-to-face contexts. However, despite the fact that this information gets exchanged, it does not get used (Dennis 1996). Thus, my primary design goal is to develop a system that addresses the latter problem.

As described in Chapters 2 and 3, the design of the platform will illustrate the two features that are the focus of this dissertation:

- The system will use information made available in structured communication channels (referred to in Chapter 3 as a Coordinating Representation, or CR) to drive adaptive functionality.
- The adaptive solution implemented will demonstrate how this information can be used to fundamentally alter the nature of the collaborative process in order to overcome the information processing problems described in the previous section.

As described in the previous chapter, the CR that will be used will be based upon prior work with argument visualization platforms, and in order to guide the group towards a a more “rational” outcome, the underlying adaptive algorithm will employ Dempster-Shafer theory to determine the best alternative given the available evidence, and use this information to guide collaborators. In the following sections, I will describe an approach for doing this.

This remainder of this chapter is divided into two main sections. In the first section, I provide an outline of the design rationale. This rationale is framed in terms of loose methodology for developing adaptive groupware, borrowing from Browne et al.'s (1990) MAID approach to the design of adaptive systems.

Following a discussion of the design rationale, I will describe the implemented system. I will not dwell upon technical details, but will provide a brief functional description as required to understand the performance of the system and a detailed description of the interface.

1 Design

There are few available methodologies that can be used to guide the design of an adaptive system. Browne et al. (1990) offer one approach, which they refer to as MAID (Methodology for Adaptive Interface Design). This methodology is based upon the specification of six "metrics," designed to help developers clarify their objectives, and focus their attention upon critical considerations. I will not adopt the entire MAID framework here, but borrow the following three metrics to help structure the design rationale presented here:

1. **Objective Metric:** The overall objective of the adaptive system. It may be to speed the user's performance, reduce errors, or increase satisfaction.

2. **Trigger Metric:** The "trigger" at runtime that should result in the performance of the adaptation. For instance, an adaptive help environment might monitor for user errors, and display a help dialog when an error is detected.

3. **Recommendation Metric:** What the system "does" in response to a trigger event. In an adaptive help system, this might be to display a specific help screen in response to a particular kind of error.

In addition to these metrics, I will also describe the mediating structure that is to be incorporated into the system, and the associated algorithm. The order of presentation in the chapter is for expository purposes only; in practice, it is likely that the design of mediating structure and the design of adaptive functionality will iterate, but the specifics of any particular solution may vary.

1.1 Objective Metric

As discussed, the common knowledge problem is robust and persistent, and can be readily evaluated though the hidden profile experimental protocol. To review, a hidden profile experiment is one in which information is distributed asymmetrically to a group of decision makers. There is a “correct” answer that can only be determined if all users contribute all of their uniquely held information. A hidden profile experiment is arranged in such a way that it is also possible to determine which answer should be chosen based on the information exchanged by collaborators.

The research community has developed a sufficient body of expertise that should allow us to replicate the problem in an experimental context, and so demonstrate a solution. However, the common knowledge problem describes a couple of phenomena and may actually have multiple causes so it is important to be precise about which of these phenomena we are interested in.

The biased sampling problem is one potential cause for the common knowledge problem. Biased sampling describes the tendency of collaborating decision makers not to retrieve the unshared information that is necessary to solve a hidden profile problem. However, previous work with GDSS supported group decision making suggests that GDSS is able to address this problem by increasing the overall volume of information exchanged by a group. For this reason, the biased sampling problem is not the focus of this thesis.

Instead I will focus upon decision quality as measured by a group's ability to combine the information they do exchange to produce a group decision that is consistent with this information as measured by a rational model. Hence my objective metric is decision quality as measured by the consistency of the group's decision with information exchanged, not as measured by the group's ability to attain an optimal solution.

1.2 Mediating Structure

In Chapter 4, I covered two argument visualization platforms, Reason!Able and Belvedere, and an argument evaluation strategy attributed to Das et al. (1997). All of these are loosely based upon Toulmin's theory of argumentation (Toulmin 1958), and each simplified that formalism in roughly the same manner. The definition and correlation between each of these formalisms is shown in Table 7.

Fox & Das (2000)	Belvedere	Reason!Able
Alternative	Hypothesis	Claim
Argument	Data	Argument
support (X)	Consistent with	Reason
support (not X)	Inconsistent	Objection

Table 7: Mapping between various argument formalisms

Empirical work (e.g. Suthers et al. 2008; van Gelder 2003) has shown that both Reason!Able and Belvedere meet the requirements for useful mediating structure described in Chapter 3. That is, these representations improve collaborative decision making along several dimensions (e.g. information integration, retention, and consensus formation), and users like to use them. Furthermore, there is a prior developed algorithmic approach to decision making that maps directly into this structure. Thus, adopting an interface structure that is similar to previously explored argument visualization platforms will produce the information required to support adaptive functionality, without interfering with users' ability to coordinate their deliberative process.

I will adopt the terms “alternative” and “argument” to describe the same elements in Das' (2005) lexicon, but will use the terms “pro” and “con” to indicate “support(X)” and “support(not X).” As with Belvedere and Reason!Able, the interface formalism will support hierarchical arguments. I will also use graph terminology to refer to topological entities; alternatives and arguments are both *nodes*, and pro and con relationships are *links*. Links are oriented in the direction of support, so an argument is the *source* of a link, and an alternative is the *target*. The entire abstract model is captured in Figure 24.

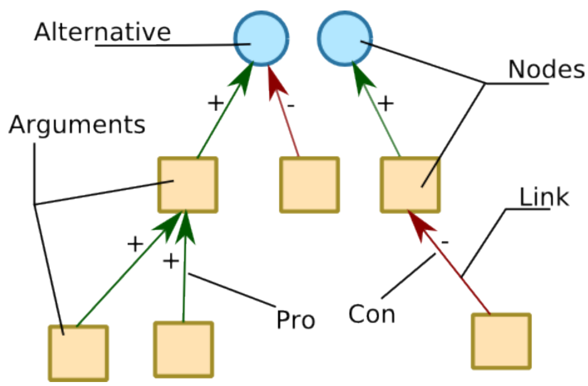


Figure 24: The anticipated mediating structure.

The described interface structure provides the runtime system with roughly the same information that is available in standard GDSS systems. With the incorporation of an algorithm based on that described by Das (2005), and a mathematically grounded belief aggregation procedure, the system will be able to determine the “winning” alternative according to a rationalistic information processing perspective. It will not, however, be able to determine if the semantic content of any particular individual contribution is valid.

1.3 The Algorithm

As discussed in the previous chapter, Das (2005) illustrates that it possible to instantiate the algorithmic argumentation framework discussed in Das et al. (1997) using Dempster-

Shafer theory. However, that framework does not explicitly cover chaining of arguments, as is required in this domain. To support chaining, we might consider any given argument to function as if it it were an alternative with respect to the arguments beneath it. However, this is not straightforward because there is no way to represent disbelief in singleton sets, and arguments as described within the formalism are singleton sets.

To handle this case without extending DS theory, we may simply pair each argument with its antithesis. These antithetical arguments would not be displayed to the user, as they would only serve an algorithmic purpose. Thus, a "con" link from argument *S* to argument *T* may be displayed as such in the interface, but would be considered by the aggregation algorithm to be a "pro" link from *S* to ~*T* (the antithesis of T). This simple extension is sufficient to support chaining.

Together, the mediating structure described above, and the algorithm described here, define the core engine that will be used to drive adaptive support in the system. They embody a rationalistic ideal, based on a mathematical model of likelihood. The following sections describe how this core engine might be used to guide users to make better decisions with respect to this ideal.

1.4 Trigger Metric

A "trigger metric," in terms of the MAID methodology, is an event or period when an adaptive support will perform its function. In Goodman et al. (2005), for instance, the trigger for adaptive support is the detection of a "bad" collaboration segment. In this domain, analogous events might be considered to be be phases or instances of "incorrect" information processing. However, it is unclear that there is any theoretical support for the existence of observable phases like this, and given the kind of runtime information made available by the mediating structure described above, it is not likely we could detect them.

There are, however, two logically distinct phases of activity in a consensus decision process: everything leading up to the final decision, and the final decision itself. I will refer to these as the deliberation and decision phases, respectively. Different sorts of adaptive functionality might be used during either phase to help guide the group to make a decision based on information that is exchanged. In order to detect the transition between deliberation and decision phases, we might simply include a "voting" tool that will help people with the bookkeeping necessary in making a consensus decision. This is a standard tool in GDSS systems (DeSanctis and Gallupe 1987), and is an example of a coordinating representation that has been developed to solve a recurring coordination problem in group decision making (see Chapter 3 for more discussion of this).

1.5 Performance Metric

Although we cannot detect actual instances of biased evaluation during the deliberation phase, the system can provide continuous feedback about the "best" alternative according to the belief aggregation algorithm based on the information provided by the users. Whenever an individual provides new information, the system can update its assessment in order to show the individual what impact their information has had on the overall deliberation. If the system assessment matches that of an individual, that individual's opinion should be reinforced. However, if the system's assessment contradicts an individual's assessment, she has several options:

- She could change her opinion to match the system's.
- She could continue to exchange information to change the system's assessment.
- She could ignore the system's assessment.

If the system's assessment mechanism is correct, each user's assessment of information is correct, and each individual user responds to the system's feedback according to the first option, we will approach a more “rational” solution. If each user chooses the second option, the outcome will depend upon the group's ability to appropriately weight each piece of information within the formalism. If, however, users ultimately choose the third option, the adaptive feedback mechanism will have had no effect.

To avoid the third option we can enforce a policy that any proposed solution must be compatible with the system's assessment. This is the second piece of the performance function, and will be triggered by the users' use of the incorporated voting tool. When it comes time for users to vote, the system might simply provide the group with a proposal matching its current assessment of exchanged information. This is not as draconian as it may sound; it guarantees that the users consider the system's assessment, but the group still needs to reach consensus around the solution, and may continue to post new information in order to adjust the system's assessment.

1.6 Summary

The preceding sections describe the envisioned functionality for the adaptive system. Informal validation of the framework was performed, and the system was found to function as anticipated[2] – that is, the system reliably identifies the correct option based on evidence for a wide range of argument network arrangements. There are, however, several anticipated failure modes in which a group equipped with this system will not produce an outcome that is consistent with relevant information in a hidden profile experiment, as follows:

- Incorrect formalism use – Users might be unable to use the formalism to express the valence of their arguments correctly, or use dialog moves which cannot be interpreted as arguments (e.g. information seeking questions).
- Incorrect information interpretation – Users might not interpret or weight information correctly.
- Too much irrelevant discussion – Users might dwell on information which is not relevant to the hidden profile, and this could sway the system away from a consistent answer.
- Gaming the system – The users might "gang" up on the system in order to achieve an answer which is not the system's assessment of information.

These are areas that will be the subject of investigation in the case study following this chapter. In the following section, I will describe the implemented system.

2 The REASON Platform

The system described in the previous section was initially implemented as a collaborative, standalone application by myself and several other developers at Charles

2 More formal studies are possible, and are opportunities for future investigation. These are discussed further in Chapter 8.

River Analytics, Inc. under a Phase I DARPA SBIR grant as a tactical decision aid. The platform was demonstrated at DarpaTech 2005, and a Phase II grant to continue development began in June 2006. Through subsequent efforts by myself and others, this original platform has been transformed into a web application designed, in part, to overcome poor collaborative information processing.

The REASON (Rapid Evidence Aggregation Supporting Optimal Negotiation) platform currently runs as a web application. Some of the implementation details that impact the functionality of the application are described in the following section.

2.1 Architecture

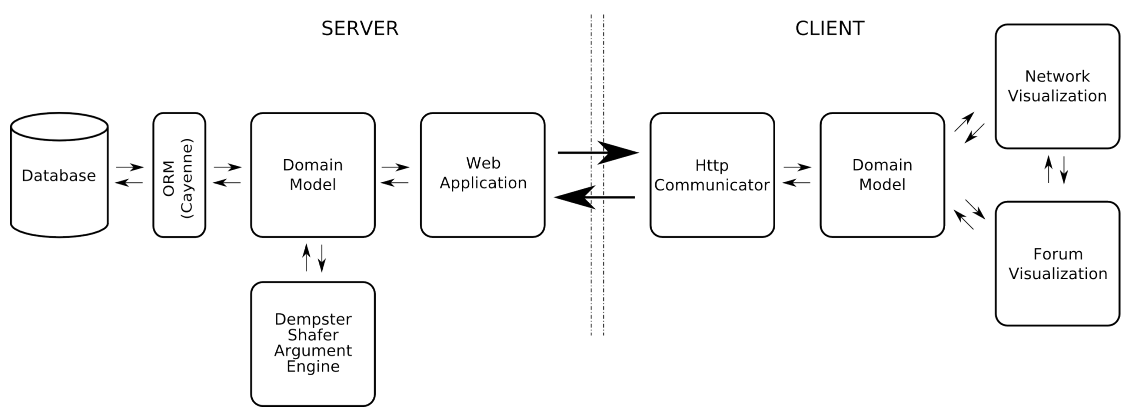


Figure 25: Functional architecture of the web application

Figure 25 illustrates the functional architecture of the system. The server consists of five main conceptual pieces. The web application itself handles all incoming requests, which read from or write to an in-memory domain model, which is a runtime model of all of the information relevant to the argument network. At runtime, the domain model itself is a cached version of information that is maintained in the database. Any modifications to the data in the domain model cause the Dempster-Shafer argument engine to re-aggregate all information and compute new weights. The logic that maps the domain model to the database is handled via an object-relational mapping (ORM) layer, implemented using an open source product called Cayenne (http://www.cayenne.org).

The domain model plays a central role for both the server and the client. This model is a taxonomic view of the "deliberation" domain, and is the formalism that was introduced in the previous section (Figure 24). According to this model, a deliberation consists of alternatives and arguments, evidence (potentially from multiple sources) and beliefs. An argument may be "pro" (for) or "con" (against) its parent, and can have only one parent. A parent of an argument is either an alternative or an argument, and an alternative cannot have a parent. Each alternative belongs to exactly one deliberation, but a deliberation can have multiple alternatives. The definition of such a model is a useful way to control complexity in design, and is one of the primary tenents of domain-driven design (Evans, 2004).

The client runs as a Java™ applet embedded in a web browser, which communicates with the web application via standard HTTP POST and GET commands over the port being used by the web application. The applet polls the web application at regular (five second) intervals for any changes that have been made to the server domain model, and propagates these changes to a local client-side domain model. However, because modifications to the server-side model that result in re-computation may incur some processing time, any given post may take somewhat more than five seconds to propagate to other clients. In practice, lag time between a post and its appearance on other collaborators' screens was never more than ten seconds.

Interaction between the user and the client-side domain model is mediated by a graph-based visualization that portrays the argument network based domain model. Interaction with the interface is described in the following section.

2.2 The User Interface

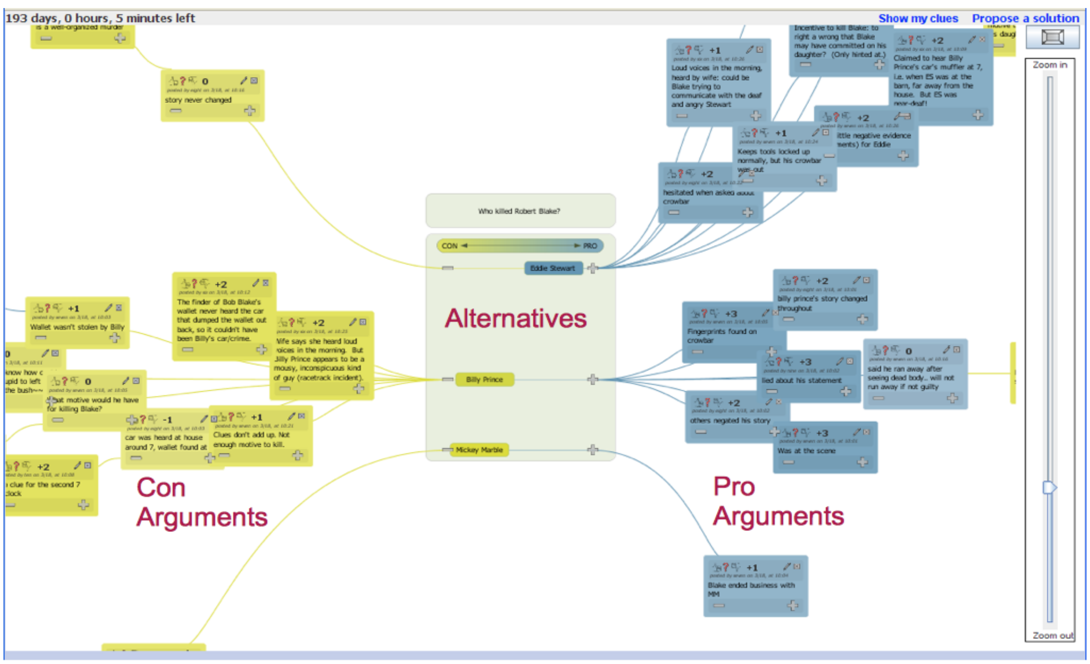


Figure 26: The REASON web interface

The interface presents a graph-based visualization of the domain model (Figure 26). The interface is built using the Prefuse (http://prefuse.org) graphing library, an open-source toolkit. Alternatives are represented by bubbles in the box in the center of the display. Arguments are represented by nodes in subtrees that are attached to the alternative. Arguments are colored yellow if they disagree with their parent, and blue if they agree. The initial argument for or against an alternative determines whether a subtree extends to the right (pro) or left (con) of the alternative. The user can zoom in and out, pan the display, and automatically re-center and fit the graph to the window by using provided controls.

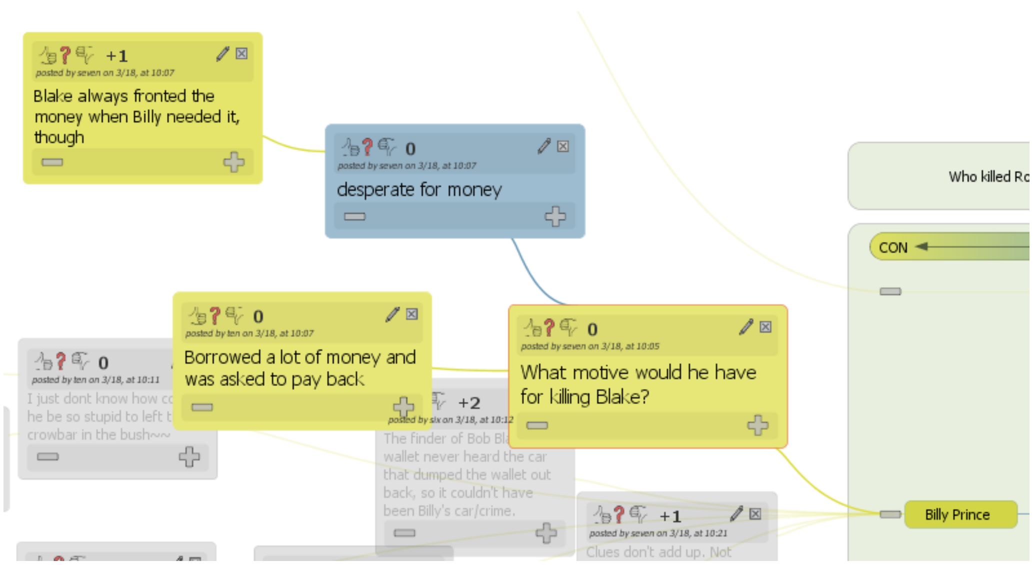


Figure 27: Selecting a node; the node with the orange outline is selected.

The nodes of the graph in the visualization "float," and are automatically laid out via an animated force-directed layout algorithm, such that nodes exert a repulsive force, and links exert a spring force. Force planes are employed to keep nodes in different subtrees separate. Because the graph is continuously animated, users are able to drag individual nodes, and this will pull attached nodes along with the dragged node – upon releasing the node, all nodes will drift back to a position determined by the layout algorithm. Clicking on any given node will select that node (see Figure 27), its ancestors (up to an alternative), and its descendants. Selected nodes are zoomed in, and the animation for those nodes is paused, so that the user can control the placement of the nodes. Other nodes (which are not part of the selected set) are deemphasized, and remain animated.

All message posting and voting occurs via the graph. Controls are included in each existing argument and for each alternative to allow the user to post a new argument for or against that item (see Figure 28). Once an argument has been posted, the user can change their vote (in favor, against, or neutral). The username of the posting user, the time when the post was made, the user's current vote, and the sum of all votes assigned to that argument are all displayed (note that because votes may be either "up" or "down" this number can be negative). The user may also edit the text or valence of the post, or delete it. Deletion will cascade to all children. Most commands (voting, posting, editing) are reflected in the interface within the update interval. Deletion happens immediately (after a confirmation dialog).

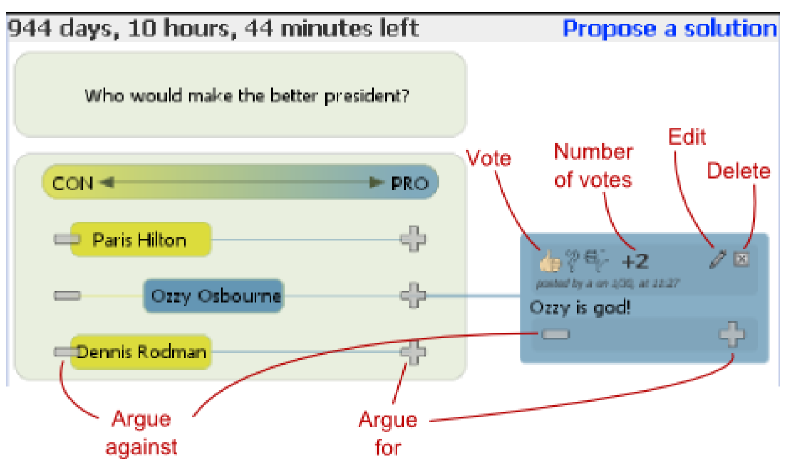


Figure 28: Node detail

The position of each alternative in the central box in the graph indicates the current aggregate belief for any alternative; as an alternative accumulates more belief, it moves further to the right, and its color changes from yellow to blue. The aggregate belief is updated as each post is made, deleted, or votes change, so there is immediate feedback for the user upon posting an argument.

2.3 Mediating the Decision Process

The interface described thus far provides one part of the adaptive functionality described in the previous section; users are provided with a continuous display of the "winning" alternative according to the belief aggregation algorithm. Several additional features mediate the consensus formation process for the groups.

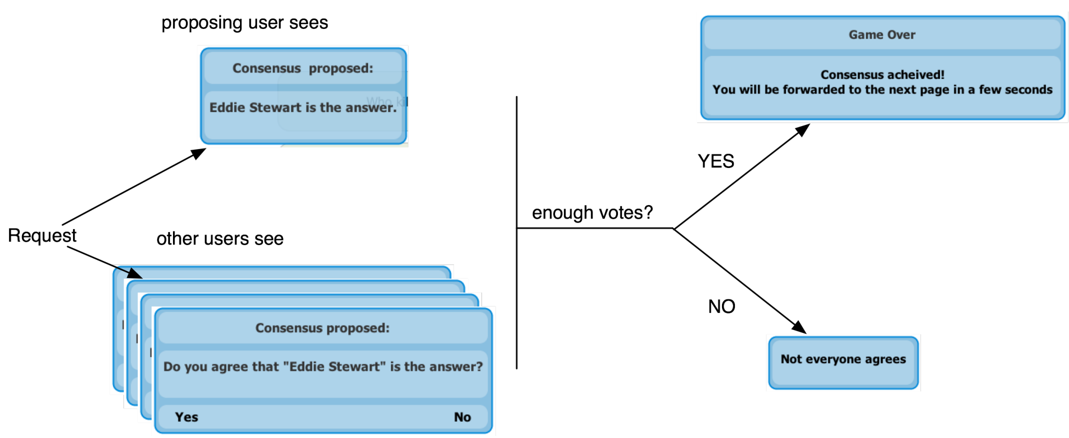


Figure 29: The voting process in REASON.

When any user feels the group is done, or wishes to propose a solution, they may click the "Propose a solution" link, which is displayed at the top of the interface. This initiates the decision process, which is structured via the set of dialogs shown in Figure 29. The system does not ask the user for their proposed solution; the current alternative with the most aggregate belief is considered to be the proposal. Thus, the group is only allowed to make a final decision that is consistent with the algorithm's assessment of the evidence.

When one user has initiated a "Done" proposal, dialogs soliciting other users' assent or dissent are displayed in their interfaces. While the request is pending, a dialog is displayed in the originating user's window, containing information about the number of assenting votes. If any of the users disagree with the proposed ordering, the decision process is canceled and all users are notified that not everyone agreed. If all users in a decision making group agree with the proposal, they are notified, and the deliberation is locked so that no further modifications can be made.

The following two chapters describe the results of a case study with the REASON platform. In the next chapter, I provide results demonstrating that the platform had the intended effects on group-decision making. However, the behavior of groups that did not have the adaptive functionality did not use the platform quite as expected. These groups did not make group decisions that were *consistent* with exchanged information, but they did make the *correct* decision nearly as often as groups with the adaptive platform. This is the subject of an extended case study in Chapter 7.

Chapter 5: A Case Study in Mediated Decision Making

The previous chapter described the design rationale for and implementation of an adaptive group decision making platform called REASON (Rapid Evidence Aggregation Supporting Optimal Negotiation). In this chapter, I describe a hidden profile experiment that was run with the implemented system to evaluate the effectiveness of the incorporated adaptive mediation. As described in the previous chapter, the intent of the adaptive mediation was to improve information pooling according to a rationalistic ideal. This ideal is based upon standard SEU maximization approaches to decision making, and suggests that each information item confers some quantity of support upon at least one of several decision options, and that these quantities may be combined to arrive at the best option.

The adaptive mediation performed as intended, and in fact transformed the collaborative process into something closer to the rationalist ideal. This chapter, and the one that follows, offer two forms of support for this claim, and should be considered to be two parts of a complete analysis.

This chapter will emphasize an information processing perspective which is informed by the rationalistic ideal. To this end, I will evaluate the performance of groups with and without adaptive mediation first by determining the aggregate "weight" of all pieces of information exchanged by collaborators, and comparing this to their actual decision. I will show that this aggregate weight was a far better predictor of the outcome in cases with adaptive mediation.

To determine the role the platform played in the decision process, I will examine how the portions of conversation that were "about" (and not about) pieces of information were represented within the formalism provided, and how the engine evaluated these portions of conversation. I will show that, in the adaptive cases, collaborators usually used the formalism properly, and that improper use or overall poor information processing led to results that were not consistent with the rational ideal.

However, there are two aspects of these results which will require an extended analysis in the subsequent chapter. First, although groups in the adaptive condition made decisions that were more consistent with their exchanged information, there was no difference between the *accuracy* of the outcome between the two conditions. Furthermore, the analytical approach used to examine portions of conversations in the adaptive case yields highly inconsistent and very hard to interpret results in the non-adaptive case.

This chapter is organized as follows. I will first introduce the non-adaptive platform, and then the design of the case study. I will then present the results of the study, and perform the analysis described above for the adaptive groups. Following this analysis, I will offer some thoughts based on my results for future extensions to the platform.

I will then turn my attention to the non-adaptive groups briefly, to show that the analysis techniques used in the adaptive case do not yield much insight. Finally, and before moving onto an extended analysis of the non-adaptive groups in the following chapter, I will offer a brief summary of the results presented in this one.

1 The Non-adaptive platform

As described previously, the goal of the adaptive mediation in REASON is to improve information pooling by a group of decision makers, such that the outcome of the decision reflects the information exchanged by the decision makers. To study the effect of the adaptive functionality, it was necessary implement a version of the system without the adaptive functionality. A screen-shot of the non-adaptive interface is shown Figure 30.

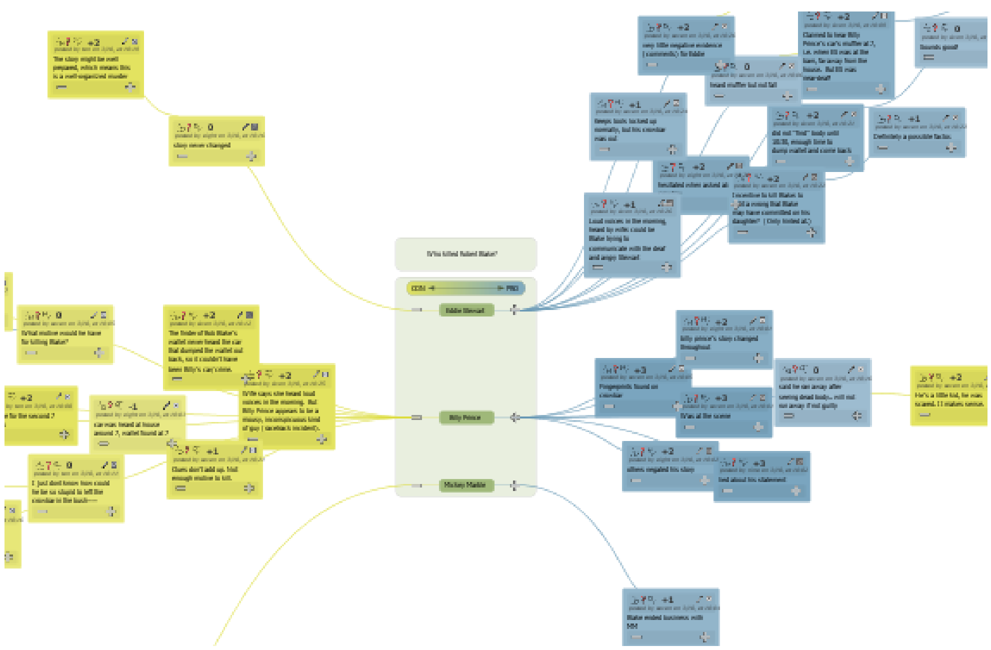


Figure 30: The non-adaptive version of REASON

The non-adaptive platform is identical the adaptive platform with two important exceptions. First, aggregate weights for the alternatives is not displayed; all alternatives are locked in the middle of the central alternative display area. Users may still vote on individual arguments, and the accumulated totals are displayed in each argument, but this has no other effect.

The decision process is also modified so that when a user proposes a vote, she is queried for her proposed alternative instead of the system supplying one. The dialog presented to the user is populated via a drop down menus, and she must select an option to submit a proposal. The subsequent decision process is identical to that in the adaptive version.

2 Case Study

The case study was based upon the hidden profile experiment described in Stasser and Stewart (1992). The hidden profile is a murder mystery, in which participants are required to exchange clues in order to identify the most likely suspect. The mystery is a slightly modified version of the the one used in Stasser and Stewart (1992).

All information about the mystery is contained in a set of affidavits and several other informational documents that are provided to the participants. The complete packet of information, and all associated clues are provided in Appendix A. Embedded in the affidavits are twenty-four clues, each of which either implicate or exonerate one of the three suspects. The suspects were E (Eddie), B (Billy), and M (Mickey), and E was the correct answer. With one exception, the clues are the same as those identified by Stasser and Stewart (1992). The clues are integrated into the affidavits, and are not distinguished from other text.

Groups of five were used in the experiment. This number was chosen as representative of small group decision making in "real-world" scenarios, and as the smallest size group which might begin to see significant process gains due to GDSS support (Fjermestad and Hiltz 1998).

Information was distributed among the participants so that three of them had some information that was unshared, and the other two had only shared information. Unshared information was distributed following Stasser, Stewart, and Wittenbaum (1995) such that each participant with unshared information had all of the unshared clues about one of the three suspects. For the remainder of the discussion, the participants will be referred to as EE (E expert), BE (B expert), ME (M expert), S1 (Shared information 1) and S2 (Shared information 2).

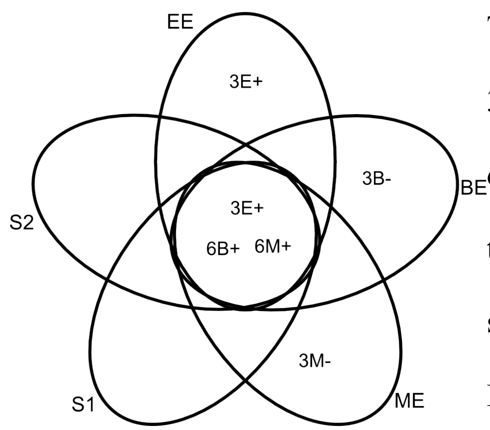


Figure 31: Distribution of information for experiment; five participants share 15 pieces of information about 3 options (A,B,C). Shared information supports either option B or C. Three participants have 9 pieces of unique information between them. All information taken together supports option A.

The distribution of information is shown in Figure 31. This distribution implies the following pre-discussion preferences: EE should be equally likely to choose any of the three options (E, B, or M); BE should choose option M; ME should choose option B, and; S1 and S2 should both be equally likely to choose either B or M. Hence, there is no statistical pre-discussion bias supporting any single option, but there is a bias away from the correct option (E).

As described in Chapter 4, prior research demonstrates that groups that are told information is distributed unevenly are better at information sharing than groups that are not, but that problems in information pooling remain. In this experiment, I was primarily interested examining how collaborators pool information, rather than how they share that information. Consequently, participants in this experiment were told that information was distributed unevenly, and that they would need to exchange their information to identify the correct option. For similar reasons, participants were also given access to their information during the collaboration.

A pilot experiment was run to validate the feasibility of the study. The pilot was run asynchronously over the course of a week with forty people, in eight groups of five. All communication with participants took place via email and phone, and all training materials were provided as help documents that were accessible from the application. As as result of this pilot study, several modifications to the experimental design were made. Most importantly, the full experiment was run synchronously, because participation in the asynchronous case was very uneven. A consequence of uneven participation was that six of the eight groups were unable to achieve consensus. In the full experiment consensus decisions were further encouraged by promising a movie ticket to all participants if they picked the correct suspect. Additionally, a training period was introduced in the full experiment so that each group had some hands on experience with the application before initiating the decision process. Finally, one of the clues in the mystery was changed because it provided an “air tight” alibi for the M alternative, leading very few groups to actually consider alternative M as a viable option in the pilot study. Both the original and modified clues are both included in Appendix A. The full experiment is documented in detail in the method section below.

2.1 Method

2.1.1 Overview

Groups of five participated in a single-factor, synchronous, web-based group decision making experiment. The decision making task was a murder mystery that required users to unanimously agree upon a guilty suspect within the alloted time. Twenty-four clues were unevenly distributed among the five members of a group, as described above. There were two experimental conditions. In the "adaptive" condition (henceforth the "A" condition), subjects collaborated via the adaptive version of the REASON platform described in Chapter 5. In the "non-adaptive" condition (henceforth the "NA" condition), subjects used the version of REASON without any belief aggregation, described above.

2.1.2 Subjects

115 university students and community members were recruited for the study. Participants self-selected into groups of five; scheduling was accomplished via a web-based application, which allowed potential subjects to choose from available time slots. As each slot was filled, an email was generated to confirm scheduled participants' availability. Any subsequent coordination problems were handled by the experimenter. Groups were assigned to experimental conditions in alternating fashion as they became available.

Each participant was paid ten dollars an hour for their time, and promised a free movie ticket if they chose the correct suspect. Subjects were not told if they had "won" until after all 115 participants had been run through the experiment.

Of the 23 groups, three are excluded from analysis due to technical problems which either interrupted the experiment or hindered communication. In total, the experimental data covers one hundred participants in twenty groups, evenly divided across the two conditions.

2.1.3 Procedure

The experiment was run synchronously in a university computer lab. Each participant was placed at a capable workstation (running either Ubuntu Linux or Macintosh OS/X) with a large format monitor. Participants were in general able to see one another, but were not able to see one another's monitors without effort. Participants were not allowed to talk to one another during the collaborative portion of the experiment, and were assigned anonymous names.

The entire experiment was implemented as a timed web-application. The phases of the experiment included an introduction, a consent form, a timed training application with instructions, a timed pre-study period, a timed collaboration period, and an exit interview. Each timed period was assigned a deadline, and deadlines for all but two of the groups[3] were 30 minutes apart. Thus, if all participants finished the pre-study early, additional time was available for the collaborative portion of the task. All participants were required to finish with the pre-study before the collaborative task could begin. In practice, most groups took the entire time allotted to each of the timed periods. All instructions given to the participants are attached in Appendix A.

The training application was identical to the application that would be used during the

3 Initial groups were given an hour for the collaborative portion; this appeared to be more time then needed, and hampered recruitment efforts. Subsequently, time was reduced.

collaborative period, but an unrelated decision problem was used ("Who would make the best president?"). Instructions were included on the same page as the training application, and there were slight differences in the text between the two conditions, corresponding to differences in the application. For example, instructions regarding voting differed according to the effect of voting. In the A groups, instructions were worded, "Vote using 'thumbs up' or 'thumbs down' to change which answer is winning." In the NA groups, the same instructions were worded, "Vote using 'thumbs up' or 'thumbs down' to express your agreement (or disagreement) with an argument." The consensus formation process was also explained during the training application, and participants were required to achieve consensus to exit the training problem before the deadline. All groups successfully achieved consensus before the deadline for the training period.

Following the training application, participants were given until the next deadline to read through their mystery materials. They were told they did not need to memorize the information, as it was to be provided during the collaborative portion of the task. Participants were required to indicate their opinion in a web form following the pre-study period before moving on to the next phase of the experiment.

Next, participants read a page with some additional information and guidelines for using the software. The page informed them that information was distributed unevenly, and that they would need to pool their information to solve the mystery. They were also reminded of some of the advice offered during the training period. This advice suggested that they post short statements that did not include compound words like "and," or," or "if;" that new clues should be posted at the "top" level, next to the alternative they addressed and that discussion of these clues should occur in the threads attached to these clues; finally, they were advised that clues about a given alternative should only be posted in the thread attached to that alternative.

After reading these guidelines, subjects were forwarded to the main application. The application displayed a countdown timer to let participants know how much time remained in the experiment. A link to their evidence set was also provided; clicking on this link displayed the evidence in a pop-up window. The collaborative portion of the task ended when either a group formed consensus or the application timed out. All but three of the groups – two in the NA condition, and one in the A condition – were able to form a consensus decision. Following the collaborative portion of the application, participants were asked to fill out a brief exit survey, which included questions about their personal opinion regarding the guilty suspect, and usability information about the application itself.

In addition to pre-discussion opinions and exit interviews, a variety of information was collected throughout the study. Each deliberation was stored in a database for subsequent analysis. All data from submitted forms (consent, pre-discussion opinions, and exit survey) was also collected in a database. A substantial amount of information was also collected in the web log. Some of the information in the web log is redundant with the database information, but other information, such as selection activities in the interface and all timing information is only available in the web log.

In the following sections, I report on the results of the case study, and offer a first level analysis of these result. This analysis is based upon those typically performed for hidden profile experiments (e.g. Stasser and Stewart 1992; Dennis 1996), and considers each piece of information to be a quantity that may be combined with other pieces of information to arrive at an aggregate weight indicating the best decision option. I will demonstrate that the adaptive application performs as intended and that groups with the adaptive application are more likely to make decisions that are consistent with the clues they exchanged than the those without. I will also examine observed and potential failure modes for the adaptive platform en route to improving its design.

Following this initial analysis of the adaptive condition, I will illustrate that this analytical approach is not as fruitful for the non-adaptive case. The chapter following this one will look more carefully at conversational content and decision procedures in order to explain the relative effectiveness of groups that did not have the adaptive platform.

2.2 The Information Pooling Analysis

The twenty groups generated a total of 1146 posts (arguments), split almost evenly between the two conditions (604 for the NA groups, and 542 for the A groups). On average, A groups authored 60.2 posts (SD=12.8) and NA groups authored 67.1 posts (SD=18.9). This difference was not significant. The entire corpus is roughly 14k words.

2.2.1 Prediscussion Choices

Despite efforts to balance the fact pattern (via the replacement of one clue, as discussed above), an examination of pre-discussion choices revealed that the information contained in the mystery was biased away from suspect M (see Table 8 for overall results). Similar results are not reported in any of the prior work using this mystery (Stasser and Stewart 1992; Stasser, Stewart, and Wittenbaum 1995; Stewart and Stasser 1998), so there is no basis for comparison.

Nonetheless, the evidence manipulation did have an impact on individual pre-discussion choices consistent with the intended impact, although there was a weaker preference for M than expected for all participants. A two-factor ANOVA over experimental condition and evidence set revealed significant differences between the evidence sets ($F(2,24)=5.23$, $p<.02$), but no differences between the experimental conditions, and no interactions between experimental condition and evidence set.

Participant	B	E	M
EE	.4	.45	.15
BE	.35	.4	.25
ME	.7	.2	.1
S1+S2	.675	.175	.15

Table 8: Distribution of pre-discussion choices for each participant

2.2.2 Discussion Content

To determine which outcome was supported by exchanged information, it was necessary to examine exchanged information for clues. The entire corpus was tagged by hand; because the posts were available in text form, there was very little ambiguity in identifying the existence of clues. A clue was considered to be in a post if the specific piece of information is mentioned, and every instance of a clue was documented for each post. The coding of clues has not yet been validated with independent coders; this is discussed in Chapter 8. For the following discussion, I will refer to the property of a clue being shared or not as the clue-class, and the suspect the clue is about as the clue-suspect.

In addition to tagging the existence of clues, several posts were of the wrong valence. That is, the valence chosen by the poster did not correspond with the content of the message. This occurred frequently after questions, which are not part of the underlying domain model. For example, in Table 9, the individual in line (2) asks whether or not there is any proof of M's alibi. This is a negative leaning question, and has been correctly posted as a rebuttal to the statement in line (1). However, while the post in line (3) would seem to agree with the argumentative force of line (2), it is incorrectly posted as a rebuttal to line (2). This is apparently a response to the first clause of the question in line (2), and so makes some sense, but is technically incorrect with respect to the argument.

Number	Valence	Post
1	[Top level - exonerates]	According to the detective's timeline, Mickey didn't have time to kill Blake and get to the golf course when he did.
2	CON	Do we have proof of when he got to the golf course, or is it hearsay?
3	CON (wrong)	I think we only have him saying it.

Table 9: Example of a post that is the wrong valence (line 3).

The above example illustrates one of the difficulties users experienced with the formalism. However, on the whole, users performed quite well considering the limitations of the formalism. Table 16 (below) contains statistics about the frequency of valence inversion. Although the rate of valence inversion is less than one in ten across all arguments, participants in the NA condition are significantly more likely to make such inversions.

Table 16 also provides several additional summary statistics for the collected data. These statistics indicate some general differences in the overall behavior of interlocutors across the two conditions. For the sake of the forgoing discussion, a thread will be defined as a chain of connected arguments, anchored at a top level alternative. There is no significant difference in the number of threads that contain clues at some point, or the number of clues per thread, but there are differences in the overall shape of the conversation. People in the NA group tend to have fewer, longer threads, and they are not as careful with choosing the correct valence for their posts. Accordingly, the number of clues per post per thread (thread clue density, in the table) is significantly lower in the NA condition. Furthermore, NA groups didn't vote as much, as is to be expected, since voting has little effect in the NA case.

	Adaptive	Non-adaptive	p-value
Average # Threads	26.4 (*sd. 8.69*)	19.7 (*sd 4.76*)	*p<.05*
Average Thread Size	2.34 (*sd 2.01*)	3.82 (*sd. 4.1*)	*p<.001*
Average # Votes Per Post	1.41 (*sd. .61*)	1.11 (*sd .31*)	*p<.001*
Average Incorrect Valence Rate (per Conversation)	.04(*sd .03*)	.10 (*sd .06*)	*p<.01*
Average # Clue Threads	.48 (*sd.5*)	.51 (*sd .5*)	--
Average # Clue Mentions / Thread	.81 (*sd .35*)	.83 (*sd .16*)	--
Average Thread Clue Density	.37 (*sd .13*)	.26 (*sd .1*)	*p<.05*

Table 10: General metrics comparing the two conditions. "Clue Threads" reflects the proportion threads that contain at least one clue; "Thread Clue Density" is the average number of clues per post per thread

Across the two conditions, participants had roughly the same level of recall for each type of clue. Figure 32 shows the relative proportions of clues mentioned by participants for each of the six categories (shared or unshared for each of the three suspects) of clues. The graph reflects the average proportion of total possible clues in each category that were mentioned at least once during a discussion. In general, unshared information is mentioned less in conversations than shared information. A three-way ANOVA (clue-class × clue-suspect × experimental condition) revealed a significant main effect for clue-class (whether a clue was shared or unshared) ($F(1,36)=6.63$, $p<.02$), but no difference across experimental conditions, and no interactions.

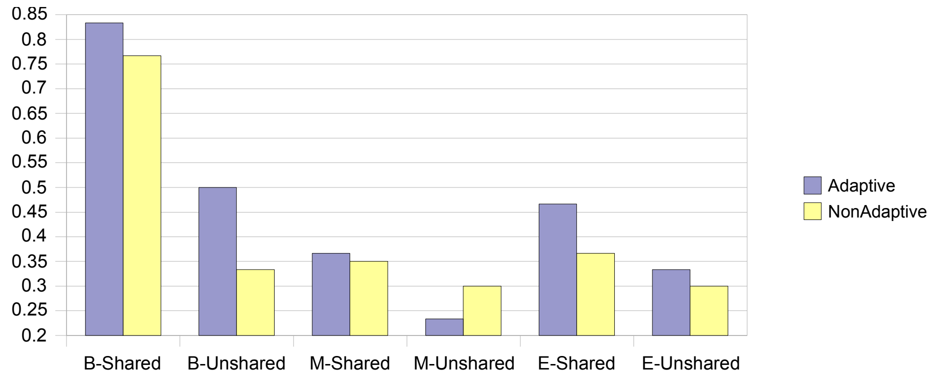


Figure 32: Shared vs. unshared clues response ratios across the two conditions

As is apparent in Figure 32 there was also a clear difference in the proportion of clues mentioned for each of the three suspects. The above analysis confirmed this, revealing a highly significant main effect for clue-suspect ($F(2,36)=8.71$, $p< .005$). This can be explained by the distribution of pre-discussion preferences. If shared and unshared clues are combined for each suspect, there is perfect correlation in each condition between the average number of clues mentioned for each suspect and overall user pre-discussion preferences ($R=1.0$ for both conditions).

This initial analysis confirms findings discussed in Chapter 4. Users recalled more information that supported their pre-discussion preferences, and universally recalled a higher proportion of shared clues.

Figure 33: Clue repeat rate

A similar analysis was performed on clue-repeat rate (the average number of times a clue was repeated). In Figure 33, the average number of times each type of clue is repeated is compared across the experimental conditions. Once again, a three-way ANOVA (clue-class × clue-suspect × experimental condition) reveals a highly significant main effect for clue-suspect ($F(2,36)=9.45$, $p<.001$). Similarly, the overall rate for each suspect correlates very highly with the pre-discussion opinions of the users ($R=.98$ for the adaptive case, and $R=1.0$ for the non-adaptive case). There is not, however, a main effect for experimental condition. There does appear to be a difference between how the two groups talk about the unshared information, but this effect is not significant.

In summary, the above analyses confirm prior findings with respect to the common knowledge problem. Participants were significantly more likely to mention shared clues than unshared clues across the two conditions. Furthermore, this effect is heavily mediated by pre-discussion opinion, which appears to dictate not only the likelihood a given clue will be mentioned, but also how frequently it is mentioned.

2.2.3 Outcomes

As discussed, the primary hypothesis established at the outset of system design was that the system would improve participants' ability to pool the information they exchanged and make decisions that are consistent with that information. Demonstration of this hypothesis indicates that the platform helps people adhere to the normative model that is implied by the representation, which based upon a Toulmin-like argument formalism and Dempster-Shafer belief aggregation, and rests upon the SEU tradition in the decision analysis literature. The results described below confirm this hypothesis.

Note that this finding does not necessarily mean that people using the adaptive platform will make the *correct* decision – factors which are not controlled for, like accuracy and completeness of information recall (which is clearly influenced by pre-discussion opinions), will still have a significant effect upon the outcome. The extended analysis following this chapter examines the relationship between what is *correct* in terms of the murder mystery, and what is *consistent* with the normative model.

Decision quality is assessed by determining whether or not the clues that were exchanged support the decision that is ultimately made by the group. Following the design employed by Stasser and Stewart (1992), each clue is given a unit of weight, and its valence (implicating or exonerating) determines its sign (+ or -, respectively). Weights for each clue that is mentioned in the discussion are then added together, and the suspect with the most weight is determined to be the best supported suspect. In case of ties, the group decision is considered to be “consistent” if it is among the best supported options.

	Adaptive	Non-Adaptive
Consistent (Unique)	4	1
Consistent (Two options)	2	0
No Decision	1	2
Inconsistent	3	7

Table 11: Decision outcomes according to clue count; the row labeled "two options" indicates that more than one answer was consistent according to the summed clues

Results of this analysis are shown in Table 11. A two-tailed, unpaired T-test reveals that the A groups were significantly more consistent with mentioned clues than the NA groups ($p<.02$). This establishes a strong correlation between decision-making behavior that is consistent with the normative rational model and the platform used. Thus, at this level of analysis, the adaptive mediation successfully transforms the collaborative process so that it is closer to the rational ideal embodied in the system.

To make the case for a causative relationship between the platform and the performance of the collaborators, we would like to be able to say that people used their available information correctly, represented it with the provided formalism as intended, and that the information represented in this manner was responsible for the final outcome[4]. Several questions should be asked. First, did conversation about clues appear to evaluate the clues correctly? If so, this indicates that participants assessed the information content of each clue as intended (validating the experiment), and that they were also able to use the formalism to encode the value of these clues (validating the platform). If clues do not correlate well with conversation devoted to clues, it is important to understand whether

4 This does not, however, imply that the platform successfully mediated the collaborative decision process. It is equally plausible that people engaged in their "normal" decision process, and did what the system wanted in order to attain their desired outcome. I will consider this question further in subsequent discussion.

the problem was with the interpretation of the clues, or with use of the formalism.

Additionally, we need to verify that it is this information that is responsible for the outcome. Does collaborators' conversation focus on clue information, or non-clue information? How are the two types of conversation related? If non-clue discussion explains the outcome better than clue discussion, it may indicate that the experimental design is flawed, but might also indicate the need for extensions to the platform to control for such “non-informational” activity. As designed, the platform assumes that people are able to focus primarily upon what is important in a decision, or at least that what is important correlates with what is not in conversation.

To answer the above questions, it is necessary to distinguish between the two sources of weight in the final decision, and evaluate them independently. To make such a segregation, I assume that any discussion following the mention of a clue in a thread is clue discussion. Table 12 provides an example of a thread that starts out as general non-clue discussion, and subsequently becomes a discussion about a clue (B's fingerprints on the crowbar).

Type	Post
Non-clue	1. (ME) I think it was either Mr. Marble or Mr. Stewart.
Non-clue	2. (BE) so how do people stand on this? Aka which one do you think is guilty?
Non-clue	3. (ME) I think Eddie is guilty but I'm still looking for a sufficient motive, probably with his daughter.
Clue (*Billy's fingerprints on crowbar*)	4. (S2) I agree with this, but I want some theory on how he got billy's prints on the crowbar.
Clue	5. (ME) Good point, I think it's possible that he removed it from the scene of the crime without thinking for some reason.

Table 12: Clue vs. non-clue discussion in a single thread. Line 4 and the following post are about the "fingerprints" clue.

For sake of this analysis then, I will define "clue discussion" as all conversation – that is, all posts in a thread – that follow the appearance of a clue in a given thread. Though such a distinction may only be approximate, it is a well-defined approach that enables automatic segregation of clue-specific from non-clue specific conversation. Inspection suggests that this is a reasonable approach in the adaptive case, as does the consistency of the data that is analyzed with this technique. This analysis is presented in the following section.

2.3 The Adaptive System

Using the above approach, the segregation of clue from non-clue discussion was done automatically, and distinct argument networks were created from these two sources of information to evaluate the weights attributable to the respective portions of the conversation. Outcomes of the resultant networks and the original network were than compared using Pearson's R to determine which portions of conversation were best able to account for the overall "weight" of the discussion. Comparisons were also made to the distribution of clues used to determine the desired outcome (according to the rational model).

Results of this analysis for the adaptive groups are shown in Table 13. In the adaptive case, by virtue of restrictions on the consensus formation process, the weight assigned to each suspect when the total discussion is considered is consistent with the final decision. There does appear to be a relevant partitioning of data with respect to the proportional influence of clue-chat; generally, groups with a positive correlation between clue-chat and the total weight of the network make a consistent decision, and those with a negative correlation do not or do not finish. Group 21 is an exception to this pattern. Furthermore, the top three groups exhibit a high correlation between the weights assigned to clue posts and the distribution of clues. This indicates that clue focused chat is roughly proportional to clue distribution, and that the argument outlined above is borne out.

Groups 12, 17 and 18 are borderline cases that do not quite fit this pattern. In the following subsections, I will discuss these three cases, and then the remaining four cases which did not achieve a correct answer.

Group #	Outcome	Group Choice	Total Weight x Clue Post Weight	Total Weight x Non-clue Post Weight	Clue Post Weight x Non-clue Post Weight	Clue Post Weight x Clue Weight
5	Consistent	E	.8	1	.84	.93
7	Consistent	B	.98	1	.99	.76
16	Consistent	B	1.0	.26	.19	1
12	Consistent	E	1.0	.41	.36	.33
17	Consistent	E	.33	.98	.12	.04
18	Consistent	E	.2	.9	-.25	.6
10	Inconsistent	E	-.98	1	-.98	.96
21	Inconsistent	M	.95	1	.94	.43
23	Inconsistent	E	-.19	1	-.25	-.96
14	No Answer	-	-.33	.41	-.99	.9

Table 13: Relative contributions of clue and non-clue chat in adaptive groups. All values are correlations (Pearson's R) between the weights for each suspect in the compared networks. "Clue Post" weight refers to weights generated just considering clue discussion; "Non-clue Post" weight refers to weights generated just considering non-clue discussion, and "Clue weight" refers to weight determined on the basis of clues alone.

2.3.1 Successful outliers

A detailed analysis of the three cases that were successful but exhibited an odd correlation patterns (groups 12, 17, and 18), suggests several different causes for these patterns.

<u>*Group 12 – Invalid Clue Interpretations*</u>

In Group 12, the relatively low correlation between clue weight and clue post weight suggests a less than perfect interpretation of the clues, although this interpretation still supports the same alternative as the clues. The group places more emphasis on suspect E than the clues (which suggest equivalent suspicion for B and E) should warrant. Examining the dialog reveals that most of the clues mentioned about B were used in a context that cast them in an opposite light (see Table 14).

Post	Clue
[BE] I don't think Billy would have killed his source of money if he has a gambling problem.	Gambling problem (should implicate B)
[EE] There was a wallet found at the car dealership without a drivers license and money. Billy doesn't know where the car dealership is.	Wallet found without money (should implicate B)
[BE] Billy lied about other things, he may have taken the wallet when he found the body.	Wallet found without money (should implicate B) Lied about being there (should implicate B)

Table 14: Incorrectly used clues for B in Group 12; all posts in the left hand column exonerate B.

The two participants responsible for the posts in Table 14 both had pre-discussion biases towards E, and this most likely led them to seek evidence to support their bias. Additionally, a third participant shared this pre-discussion bias, which may explain why none of these interpretations were countered. However, the formalism itself was used correctly in each of these cases to express the force of their respective arguments. This indicates some of the limitations of the experimental design. In a murder mystery, the valence of any piece of information with respect to a given suspect can vary depending upon the context in which it is viewed. Such a "many worlds" interpretation of information raises problems for "realistic" intellective tasks.

Group 17 – Wrong Valence Posts

In group 17, clue-post weight appears not to be well correlated with either clues or total weight. Upon examination, it was found that this is the result of a single wrong valence post (see Table 15), which gets four votes and drives the weight of the clue posts for E down substantially. Fixing this valence, and recalculating the weights restores the correlation between the weights attributable to clue-posts and the final total to .91, the correlation between clues and clue-post to .73, and increases the correlation between the clue and non-clue weight to .77.

Number	Valence	Post
1	[Top level - implicates]	His comment about hearing Billy's car is suspicious, considering his hearing issues.
2	CON (***wrong***)	Also, Billy said he likes to take good care of his car, which would mean the muffler probably wouldn't be shot. I wish we knew if that was true.
3	PRO	Also, the store owner who found Blake's wallet heard a car outside his shop, but only because the tires screeched. It wasn't because of the muffler.

Table 15: Wrong valence post that skews weight in Group 17

The error is not the only one of its kind (see, for instance, Table 9) and is indicative of the potential brittleness of the representation, especially with as limited a taxonomy of dialog moves as is currently supported. If, for instance, the platform were to include an "Extends" dialog move (as in the IBIS formalism), the mistake shown in Table 15 might not have occurred.

Group 18 – Information Processing Asymmetries

Group 18 presents a somewhat different case. Based on the data in Table 13, it would appear that the group did a reasonable job interpreting the clues (the correlation between Clue Post Weight and Clue Weight is .6), but that the non-clue discussion dominates the discussion, and is ultimately responsible for the outcome.

In general, the group exhibits very poor information retrieval. In particular, the EE participant (who has all of the unshared clues about E) contributes very little to the discussion (roughly 11% of all posts), and does not mention any clues. As a result, only one of nine threads about the E suspect are part of clue-chat. The bulk of the reasoning done by the group can be attributed to the BE participant, who is responsible for the two shared E clues that are mentioned, and is able to reason from the exonerating clues about

B to the correct answer.

Although the conversation in Group 18 about suspect E is dominated by non-clue chat (due to the dearth of information available on account of poor information recall), this chat is dominated by BE reasoning "out loud" in favor of E. Eleven of the eighteen posts in favor of E are authored by BE, and these eleven posts make up 55% of all BE's posts. Furthermore, while not specifically about "clues," they are very relevant, and capture the reasoning process that supports the selection of E (see Table Table 10 for some examples). Thus, the formalism is used correctly, and the outcome reflected by the system captures the overall reasoning process in the group accurately. This example does illustrate some of the limitations of the experimental design in that it cannot account for a good reasoning process in the absence of actual clues. I will say much more about this in subsequent sections and chapters.

Posts implicating E, authored by BE
Used gloves, used crowbar, billy touched, threw into bushes.
Alibi is not watertight, he had enough time to frame price, go back to the barn, and then come down hour later and act surprised.
Had the perfect opportunity to frame Prince.

Table 16: Posts by BE that illustrate accurate reasoning to indict E.

In summary, the three preceding cases illustrate some potential weaknesses of both the platform and the experimental design. Pre-discussion bias can lead people to misinterpret evidence (or at least, interpret evidence differently than a researcher might expect), and if a group contains enough members that share a bias, these interpretations will go unchecked. The valence of a response can be difficult for untrained users to establish, especially if the dialog move has no convenient representation within the limited taxonomy offered. Finally, asymmetrical information retrieval capabilities among group members can lead to situations where it becomes less likely that the platform will "do the right thing" with respect to the criteria established by the experimental design. Nonetheless, the platform does support a reasoning process that allows interlocutors to overcome such information deficiencies, which is indirect support for the design of the system.

Despite these difficulties, in each of these cases groups "did the right thing" with respect to the system and the experimental design. Taken along with the three cases which did not experience similar difficulties, this lends credence to the experimental design, and to the platform. In general, decision makers seem to do a pretty good job at interpreting evidence, their conversation within the framework imposed by the system corresponds with this interpretation, and the outcome provided by the system accurately reflects this information. I will now turn to the cases in which failure occurred.

2.3.2 Unsuccessful Groups

Groups failed to achieve the desired result for a variety for reasons, and each of these (as with the successful outliers above) illustrate a potential failure mode for the system. These are discussed in the following subsections.

Group 21- Gaming the system

Group twenty-one is the most intriguing case, because there were high correlations between the dialog devoted to the discussion of clues, dialog not devoted to clues and the total weight assigned. The reason that the decision was not coherent with the clues themselves is because the weight given to chat was not in fact proportional to the clues (see Table 17).

Group 21 Weights	B	E	M
Clue Weight	4	1	2
Just Clue Chat Weight	.34	.01	.56
Non Clue Chat Weight	.34	0	.36
Total Weight (Actual)	.46	0	.51

Table 17: Distribution of weights as compared with clues for Group 21

Looking closely at the data for this group, it becomes apparent that the group deliberately swung the system's assessment to M from B based on the decision they had come to in their discussion. Evidence for this comes from two sources – an inspection of the team's conversation, and an examination of their posting behavior.

It is clear from the transcript that although the system indicates that B is the best suspect, the team actually thinks it is M. Roughly seven and a half minutes before the time is up for the decision, the exchange shown in Table 18 appears.

Number	Valence	Post
1	[Top level – implicates M]	So either we all vote for prince in the next 5 minutes, or we agree to change the totals so that it says Mickey... but we have to make us and the computer agree....
2	PRO	So how do we change it to Mickey?

Table 18: Conversation for Group 21 indicating intent to "game" the system

Following this exchange, three extra votes are cast on implicating arguments for M and exonerating votes for B, four new posts appear implicating M, and the group makes a final decision with five and a half minutes remaining on the clock. Thus, it appears that Group 21 would have made a decision that is consistent with the clues, but deliberately swayed the system's assessment at the end.

Group 10 – Poor information processing and retrieval

Group ten chose E as an answer, though the distribution of clues suggested B would be more appropriate. As with group 18 above, this group exhibited very poor information recall, and this is the fundamental reason why their clue posts are not proportional to the final outcome. Out of all twenty groups, this group mentioned the least total number of clues (four implicating B, two implicating M, and one unshared implicating E). In their resultant discussion, eight of fifteen threads talked about clues for B, three of seven for M, and two of eleven for E, leading clue chat to strongly favor B. The group arrives at their conclusion by pointing out many “inconsistencies” in E's testimony which are either incorrect, or not very relevant (see Table 19 for a list of the major inconsistencies identified by the group), and which they are subsequently unable to rebut. The last line of Table 19 captures the group's final rationale for their decision. In conclusion, based on this group's performance, it would seem that one possible failure mode for the adaptive platform occurs when the group exhibits universally poor information recall and processing.

Reasons That Implicate Eddie	Assessment
Daughter left work for “personal” reasons.	*True, relevant, but does not mention associated clue. No follow up, not identified as motive.*
if you look at the first interview with mrs blake and eddie it says that mrs blake was aware of the situation at around 10:30 or so when eddie knocked on her door, but in eddie interview we see that it appears that he told her somewhere in the wee hours of the morning which shows that eddie is lying....maybe?	*incorrect*
Well it appears that eddie is all screwed up with the time.. as in times robert usually comes homes... stated that the barn is about 200-300 yards away... when it is actually 400 feet away if you look on the graph	*incorrect*
It was his crowbar even if it had billy's fingerprints.	*True but irrelevant. Reference to clue implicating B (fingerprints).*
Eddie said it rained, didn't want to get car stuck on gravel path, mrs. Blake says no car in driveway.	*True, relevant, contains unshared clue implicating E.*
Downplays what Mrs. Blake says, she says he yelled urgently, he said she was calm.	*incorrect*
Its weird that he couldn't tell that it was Blake's body.	*irrelevant*
There are so many inconsistencies... with his story... the crowbar belongs to him ......hes is confused as shit...	*SUMMARY*

Table 19: Major reasons given by Group 10 for implicating E; sorted from top-down by weight in argument network

Group 23 – Contentious disagreement

There are several factors which lead to Group 23's failure to achieve a solution that is inconsistent with exchanged clues. Table 20 summarizes all of the weights described in the following discussion.

Group 23 Weights	B	E	M
Clue Weight	5	4	3
Actual Weights			
Clue Chat Weight	.14	.24	.53
Non Clue Chat Weight	..06	.8	.08
Total Weight	.05	.81	.12
Fixed Valence			
Clue Chat Weight	.25	.35	.22
Non Clue Chat Weight	.05	.75	.14
Total Weight	.05	.72	.21
No Extra Votes + Fixed Valence			
Clue Chat Weight	.19	.1	.12
Non Clue Chat Weight	.08	.27	.1
Total Weight	.16	.28	.17

Table 20: Distribution of weights in Group 23 for several manipulations. Actual Weights are those taken from the final argument network; Fixed Valence are those with valence mistakes fixed; No Extra Votes + Fixed Weight are those with all additional votes removed for each post and valence mistakes fixed.

First, note that the clues as a whole favor B, the total weight and non-clue chat weight strongly favor E, and the clue-chat weight favors M. However, fixing the valence of one mistaken post brings the clue weights into relative alignment with the rest of the posts (R=.9). However, this distribution is still not consistent with the distribution of clues.

An examination of the posting and vote distribution in the groups conversation revealed a somewhat skewed distribution. In general, the group had many posts (the most of all adaptive groups), and a high voting rate (third highest of the adaptive groups, fourth highest overall). Furthermore, two of the users are responsible for 69% of the posts, and 83% of the 59 votes (see Table 21).

User	# of posts	% of total posts	# votes	% total votes
BE	28	.40	31	.53
S2	24	.29	18	.31
EE	18	.14	3	.05
S1	10	.10	4	.07
ME	7	.07	3	.05

Table 21: Posting and voting distributions for Group 23.

In examining the text and distribution of the posts, it is clear that the two top posters were engaged in a “war” - S2 was an advocate for M, and BE was an advocate for E. Eliminating votes and recalculating the weights of the network (the last three lines of Table 20) realigns the clue chat weight with the clues ($R=.73$), but does not change the relative ordering of support for the suspects in non-clue chat. In conclusion, the distribution of weights in this group's deliberation is due to a contentious disagreement between two users, and one of the users was able to overwhelm the other by posting more non-clue chat, and and voting more.

Group 14 – Incomplete story formation

Finally, Group 14 was unable to complete the experiment, because S1 was unwilling to agree with the system's assessment (M), which was not in fact compatible with the clues. The correlation between clues and clue-chat is very high, indicating that clues were interpreted correctly. However, unlike the other groups, neither clue nor non-clue chat correlate well with the overall weight assessed by the system. Upon examination, this is because the group is unable to recall any clues for E, yet post a relatively robust discussion (roughly one quarter of all posts) about him. In general, the group's overall recall of clues is very poor – only one unshared clue is recalled. Because the clues are interpreted correctly, clue-chat sways the outcome in one direction (implicating M and B), and non-clue chat the other (implicating E). Neither are particularly dominant, and as a result, neither correlate well with the total weight assessed by the system, or with each other.

The various manipulations – fixing incorrect valence posts and eliminating votes – reveals little. The group's indecision does not even seem to be a case of “sour grapes” resulting from pitched battle in the conversation. In the group, EE expresses the most emphatic opinion (65% of EE's posts are for B; EE is responsible for 32% of the group's posts), but is willing to vote with the group for M. S1 favors E (50% of S1's post are for E, and S1 is responsible for 15% of the group's posts) but does not seem to take a strong position in the conversation. In an informal sense, the group does not cover much territory with their discussion, and their process seems unfinished in that they never seem to arrive at a “story” that explains the clues. This may be a causal reason for the relatively low correlation between the different segments of chat and the total. I will revisit these observations in the subsequent chapter.

2.3.3 Conclusions about the Adaptive System

The usage data from the adaptive platform is a lens through which to examine both the experimental design, the platform itself, and finally, upon the ability of the platform to help people overcome problems observed in other hidden profile experiments (e.g. Dennis, 1996). The data seems to bear out that, generally speaking, both the experiment and platform perform as intended. In the case of the experiment, it seems that the weight of individual clues were generally assessed appropriately, and a simple summing of the weights is a reasonable estimate of the probable guilt inferred by groups. With regard to the platform itself, it seems that people are able to use the formalism reasonably well, with limited training, to represent the intended value of the information, and that the belief aggregation mechanism accurately combines these values to arrive at a decision that is consistent with information exchanged. Reflecting upon existing research on the common knowledge problem, it seems that the system rectifies the problem of overweighting shared information in groups of decision makers.

However, I have also identified a number of potential failure modes, which reflect both upon the experimental domain and the platform, and these illustrate some clear directions for future work. These are:

1. Incorrect Data Interpretations – As is well described in the literature (see Chapter 4), pre-formed beliefs can lead individuals to interpret information as being consistent with those beliefs. Although this is generally seen as negative, context clearly influences the interpretation of any given piece of data and it is not always an easy matter to select the correct possible world interpretation. The platform allows such confirmation bias to be corrected if there are others in the group who can correct it, but if enough people support the wrong interpretation, the platform will simply reflect the wrong interpretation.

2. Wrong Valence Posts – The problem with wrong valence posts is most likely a problem with the formalism employed. However, choosing the right taxonomy for dialog moves, as with any interface design, requires balancing usability (and simplicity) against flexibility. The platform as implemented has begun with the

simplest possible taxonomy required to support this form of belief aggregation, and it was found to be highly usable. It can be considered a baseline from which to begin investigating necessary extensions.

3. Information Retrieval Deficiencies / Asymmetries – The platform does not prevent individuals from coming forward to fill a gap in the collective knowledge of the group, perhaps by clever inference. However, it does little to proactively address such gaps. A more specific representation for a given decision problem might be able to highlight areas of missing information, much as Suthers (1999) has illustrated that matrices are effective tools for highlighting gaps in a set of information. Building a general purpose tool to achieve the same kinds of affordances is a challenge for future work.

4. Gaming the system – People are clever, and collaborators can, and will, game the system to achieve ends that are not intended. One way to deal with this might be to counter such attempts by detecting deliberate manipulations. This might be an appropriate response if there was an unchangeable incentive structure that led to such behavior. However, in this case, we would hope that the jointly constructed representation is compelling enough to participants that they feel no need to game the system.

   The fact that people did game the system indicates that the result produced by the system did reflect what the users interpreted was the overall meaning of the their conversation – even though the best choice according to the system was an accurate reflection of the pieces of information participants had exchanged. This may indicate that the system needs to be redesigned so that it is easier for collaborators to see why the system assesses the exchanged information as it does. Alternatively, manipulation of the final result could be more directly supported, allowing people to step back, evaluate their work, and work within the tool to come up with a better outcome.

5. Contentious disagreement – In my data, contentious disagreement lead to the posting of multiple non-clue posts, which biased the system in the wrong direction. An approach to addressing this problem exists within the current system, because it is possible within the existing system to manipulate the weight conferred upon any argument by a given source. It would be a simple matter condition the weight of a given to upon a set of grounding criteria. These might be explicit links to trusted sources of information, or perhaps a linguistic analysis of the "provability" of any post (e.g. Saurí and Pustejovsky 2007).

6. Incomplete story formation – Story formation is a large topic, which I deal with more completely in following chapters. However, the possibility an unclear weight distribution reflects an incomplete story highlights several avenues of further investigation. An additional analysis with the existing data should be performed to explore how the weight distribution changes over time, and whether the weight distribution "collapses" into a recognizable profile once a story has evolved. Such an analysis might further be used to reflect upon the degree to which the tool mediates group decision making process, and the degree to which people use the tool to produce a desired end-product they have achieved already

via different means. Correspondingly, if the latter occurs to a large degree, does this produce a feedback cycle that changes information seeking behavior?

I will now turn my attention to the non-adaptive platform, in order to understand how people fail to achieve results that are consistent with their exchanged clues.

2.4 The Non-Adaptive System

Group #	Clue Prediction	Group chose	System chose	Total Weight x Clue Post Weight	Total Weight x Non-clue Post Weight	Clue Post Weight x Non-clue Weight	Clue Post Weight x Clue Weight
11	B M	M	M	.32	.96	.05	.89
19	B	M	B	-.15	.36	-.98	-.53
22	B	E	B	.68	.99	.77	.96
6	B	-	B	.97	1	.96	-.96
20	E	B	E	1	.77	.7	.6
13	B	E	E	-.86	1	-.86	.88
15	B	E	E	.95	.97	.86	.31
2	B	E	E	.91	.99	.82	-.11
4	B	E	E	.89	.87	1	.43
8	B	-	E	-.3	.86	-.75	.9

Table 22: Columns 2-4 are outcomes; highlighted areas match. Remaining columns are degrees of correlation between various portions of chat (Pearson's R).

It is possible to perform an analysis of the non-adaptive groups similar to the one used for the adaptive groups. However, as I will illustrate, such an analysis may not be fruitful. As with the adaptive case, Table 22 portrays correlations between various portions of chat, as well as the group decision, the outcome supported by the clues, and the outcome supported by the belief aggregation algorithm. Because it is not clear how voting should be interpreted in the non-adaptive case, and there are significant differences in voting between the two conditions (see Table 16), all votes were removed to derive the results.

Additionally, all valence mistakes have been fixed because of the increased tendency of users to make such errors in the non-adaptive case.

Compared to the adaptive case, there is on average less correlation and higher variance in the correlation between clues and clue posts (.54; *sd=.42*, vs. .81; *sd=.16*). This result is not quite significant at the .05 level (*p=.08*). There are no other notable differences within each type of correlation (each column of numbers in the table) between the two conditions. However, unlike in the adaptive case, there is very little to suggest a partitioning of the groups via weights from the different portions of the conversation. In fact, the only noticeable pattern is the persistence with which groups seem to *avoid* a decision that is consistent with the exchanged clues. With the exception of Group 11 (which seems to perform as if it had the adaptive system), and the two groups that did not finish, groups decide against the system when the system reflects the clue distribution, or with the system when it does not.

One explanation for the apparent lack of patterns across the distribution of weights from different portions of the conversation is that the approach used to segregate clue-chat from non-clue chat was not valid. It may be that the underlying assumption – that clue-chat can be separated from non-clue chat simply by cutting each thread where a clue appears – does not hold in the non-adaptive case. There are several indicators that support this hypothesis.

As previously described (Table 16), participants in the non-adaptive case make more valence mistakes and create longer threads that are not as densely populated with clues. There are also significantly fewer threads per conversation. Figure 34 plots the length of each thread in each deliberation, sorted by length along the X-axis, colored according to experimental condition.

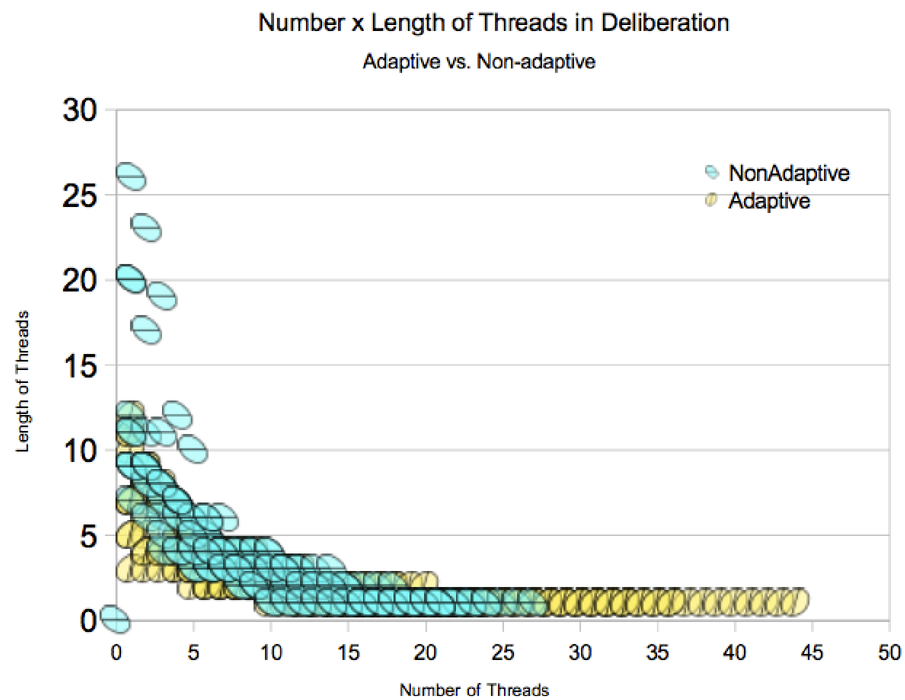


Figure 34: Length vs. number of threads in the two conditions.

In addition to differences in thread metrics, participants are significantly more likely to argue both sides of an issue in any given thread. 35% of threads in the non-adaptive case contain at least one instance of a participant posting an argument counter to their initial position, and in those threads, an average of 1.58 people will switch sides. In the adaptive case, this behavior occurs in 20% of all threads, and 1.43 people will switch sides. The difference in number of threads that exhibit this behavior is highly significant (*p<.001*), though the difference in the number of people who switch sides is not.

The above data indicate that people in the adaptive case use threads differently than do people in the non-adaptive case. In the adaptive case, collaborators create short, clue dense threads, perhaps following the instructions they received (to post new clues in new threads). In the non-adaptive case, collaborators have longer, more diplomatic conversations, perhaps covering more topics, and possibly not following their instructions quite so closely.

Because of this difference, splitting the conversation based on the assumption that a thread is “about” a clue for its remainder once a clue has appeared may not be valid. One possible approach to rectifying this problem would be to introduce a more sophisticated means for splitting up the conversation, or to tag each post by hand. However, either approach is subject to many possible sources of error. In order to gain some insight into how the non-adaptive groups made their decisions, I turn to an analytical approach that examines the content of their conversations more carefully. This analysis is described in the next chapter.

3 Summary

The results covered in the preceding sections indicate that the adaptive mediation did indeed help users to make decisions that were consistent with the information that they exchanged. In half of the cases that had a consistent answer the formalism was used as intended. Decision makers evaluated the clues that were exchanged correctly, used the formalism to represent those clues correctly, and these evaluations were properly weighed by the formalism to support a final answer that was consistent with the exchanged information.

In the other half of the cases that made a consistent decision, the formalism was used in a less than ideal manner, but these deviations were not enough to move the system's aggregate answer away from the consistent one. Furthermore, in each of these cases, the deviation resulted from a behavior that was explainable within the boundaries of my analysis. In one case, a single wrong valence post skewed the distribution of weights within the discussion. In another case, significant pre-discussion bias led users to incorrectly evaluate some of the information items. And in the final case, one user was able to overcome poor information retrieval by another user by making effective inferences, which took the form of “non-clue” discussion, but were relevant and added to the formalism correctly.

The four cases that did *not* achieve a consistent answer were also explainable within the boundaries of my analysis, further validating my results, but also indicating the need for modifications. In one case, a group achieved the wrong answer because they deliberately “gamed” the system. In another case, the group was characterized by highly contentious posting behavior indicative of pitched battle between two of the collaborators. Another group simply exhibited remarkably poor information retrieval and processing. And a final group never seemed to construct a collaborative story that was compelling enough to support a final consensus.

Each of the these failures, and the “odd” cases above, indicate paths along which the system might be extended, and I have enumerated these above. However, each of these potential failure modes also indicate that the fundamental approach was sound. People in general were able to use the system, and it mediated their interaction as intended. When errors were made, or people “abused” the system, or groups were simply not very good at the assigned task, the system failed in predictable ways.

The non-adaptive groups present a somewhat more complicated story. Only one group made a decision consistent with their retrieved information, and hence support the claim that it was indeed the adaptive mediation that led to the success of the adaptive groups. Yet the non-adaptive groups got the “correct” answer almost as frequently as the adaptive groups. They also appeared to use the platform in an entirely different manner than the adaptive groups, rendering the analysis performed for the adaptive groups ineffective. Further analysis is required, and I report on this in the next chapter.

Chapter 6: A Narrative Analysis Of Mediated Decision-Making

In the previous chapter, I examined the decision-making behavior of groups from an information processing perspective that was based upon a rationalistic ideal. From this perspective, decisions that are consistent with aggregate weight of the bits of evidence the group evaluates are the ideal. What is consistent is not necessarily the correct answer, though, because people may exhibit imperfect information retrieval and may not discuss all available evidence. My analysis revealed that groups of decision makers whose decision-making behavior was mediated by an adaptive groupware platform (the “adaptive” groups) were more likely than groups using the non-adaptive platform (the “non-adaptive” groups) to make consistent decisions. However, both groups made correct decisions about the same amount of time.

Furthermore, in the adaptive groups, it was possible to perform a surface analysis of each group's conversation in order to understand the relationship between their conversation and their outcome. This analysis revealed that the adaptive groups made consistent decisions because their conversation about evidence was well correlated with the evidence exchanged (indicating that they evaluated evidence correctly), and non-relevant conversation was either also correlated or did not have a large impact on the outcome. Where this was not the case, further investigation revealed that either misuse of the system or generally poor information processing could explain the results.

The non-adaptive groups, however, could not be analyzed in a similar manner. A surface analysis of the arguments revealed no clear patterns. I hypothesized that the reason for the apparent lack of regularity could be attributed to differences in the way the non-adaptive groups used of the platform, and that the analytical approach used for the adaptive groups could not be applied to the non-adaptive groups for this reason. In particular, non-adaptive groups tended to organize their conversation into longer threads, and were more likely to argue both sides of an issue within a thread.

In this chapter, I will perform a deeper analysis of the collected conversations in order to explain both the non-adaptive groups' surprising ability to get the “correct” answer and their somewhat different use of the formalism. This analysis will elaborate upon the differences in the decision processes used by groups in either condition, and in so doing will further support my claim that the adaptive mediation substantially changed the collaborative process of groups with the adaptive platform.

In performing this analysis it became apparent that in both conditions (adaptive and non-adaptive), the process of solving a mystery involved creating a “story” as an explanatory principle for the clues that had been identified. There is precedent for this observation in literature on decision making, though not within the literature on rationalistic decision making. Pennington and Hastie (1986) performed a series of empirical studies on juror decision making, and found that individual jurors develop cognitive story representations, which are used to mediate the verdict creation process. This work may be seen as belonging to part of the "naturalistic" school of decision making.

As an introduction to my analysis, and in order to provide support for some of the techniques I have employed here, I will first briefly discuss the naturalistic approach, and then offer a comprehensive overview of Pennington & Hastie's work. Following this introduction to the potential role of stories in decision-making, I will turn to the analysis of my data.

First, I will examine areas of *shared consensual focus* in the collected conversations, during which multiple participants converge to a common physical area of the mediating artifact to work out critical details of the story. As part of this analysis, I will introduce a quantitative approach for identifying these focus areas. The developed quantitative metric will be used as a tool to examine story creation across the two conditions.

Following this initial analysis, I will reconstruct the story creation process for two groups - one in either condition - to illustrate how pieces of data are passed through the group to become inferences, and ultimately story elements that determine the group's answer. This work is heavily informed by Pennington's (1981) work. Finally, I will use these analyses to reflect upon the data described in the previous chapter.

1 Naturalistic Decision Making

Over the last few decades, there has been a growing body of work that stands in opposition to, or at least does not embrace, SEU-based approaches to decision analysis. While each specific approach is unique, the term "naturalistic decision making" has been used to characterize these approaches, and it highlights their emphasis upon *in situ* decision making in realistic scenarios.

Lipshitz and M. S. Cohen (2005) argue that such naturalistic approaches can serve as a normative model. They observe (echoing observations by Bell, Raiffa, and Tversky 1988) that any normative theory is accepted as normative only because it meets some criteria of plausibility in the eyes of experts, philosophers, and people in general as observers of apparent truths in the world. The argument follows that the performance of expert decision makers in realistic circumstances is at least as plausible (and perhaps more so) a basis for a decision procedure as the "truths" upon which rational models are built.

A subtext to this critique is that prescriptive approaches from the "rational" perspective are based on the observation that people tend away from rational behaviors – that is, behaviors consistent with Savage's fundamental assumptions – when studied in controlled settings with contrived problems. The argument is made that this apparently irrational behavior is in fact not irrational when deployed in the context of realistic settings.

The field of naturalistic decision making is very rich and many models have been developed, although these models tend to be descriptive rather than prescriptive (see Lipshitz 1993 for a review of nine such models). I will say a little bit more about this work in Chapter 8, but for now turn my attention to one model that is particularly germane to the current investigation.

Penningon & Hastie describe a model called Explanation Based Decision Making (EBDM; Hastie and Pennington 2000), based upon extensive analyses of decision making in juries. In earlier work, Pennington & Hastie referred to what is now called EBDM as the “story” or “narrative” model of decision-making (Pennington 1981; Pennington and Hastie 1986), because they found that individuals piece together evidence to create a story before deciding upon a decision outcome. The story consists primarily of a causal and temporal model of events that describe what happened and why. Pennington and Hastie (1986) also suggest that there is an additional layer of organization on top of this, corresponding to an episodic model (see Figure 35, below), though this model is used more as a perspective from the point of view of the researcher, rather than a self-evident organization that is apparent in the decision making of individuals.

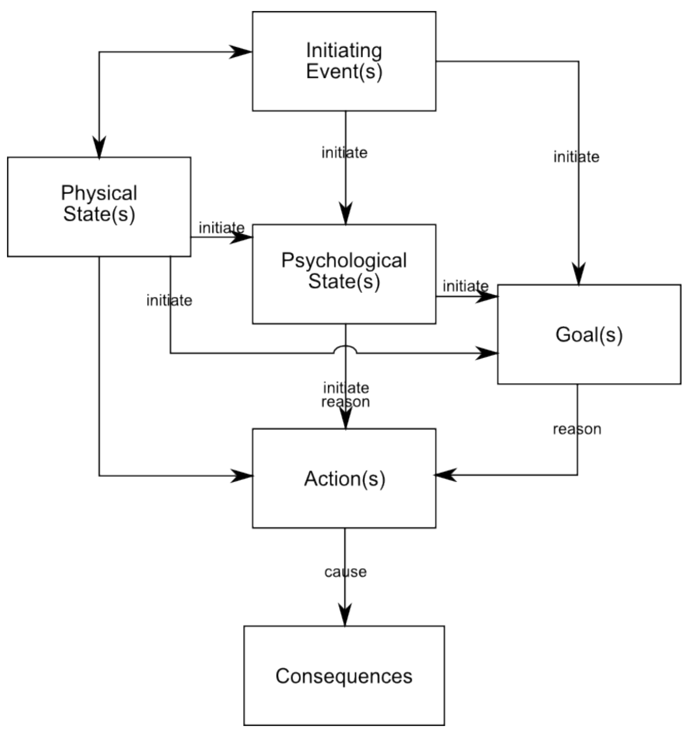


Figure 35: The episodic model used in Pennington & Hastie (1986); events may initiate state or goals; physical states may enable action and initiate goals or states; psychological states can be a reason for action, and actions result in consequences.

Pennington & Hastie argue that the story serves as an internal mediating representation, for deriving a verdict for a given case. The verdict is governed by yet another model, which establishes a set of criteria for each verdict category. In Pennington & Hastie's formulation, this model is essentially a matrix of possible verdicts and criteria within categories that define them. It is interesting to note that while Newman and Marshall's (1991) analysis of legal argument (discussed in the previous chapter) was not intended as either a representation of cognitive structures or a descriptive model of argument, their distinction between macro- and micro-level argument analysis is similar to the distinction between story creation and story evaluation proposed in the EBDM model.

To derive this model, Pennington (1981) analyzed transcripts of interviews with jury members. These transcripts were passed through a multi-level coding scheme. At the first level, interview text was coded into six "gross-content" categories, including categories for text about the story (references to events that occurred among the witnesses and suspects, as well as evaluations and hypothetical statements about the story), and references to the verdict. The story category was further broken down into four sub-categories, including: explicit story references, evaluative/hypothetical statements, "explanative" statements, and self-reflective statements. The first category dealt with "facts" as they were presented, and the later three were considered "meta-statements" about the story.

Statements in the "explicit story reference" category were then transformed into propositions and the jury member's stance (affirmed, neutral, or denied) with respect to the proposition recorded. Propositions were of eight types (act, state, mental act, goal, etc.) and were assigned a content code from a hierarchically organized taxonomy of information about the case, compiled from interviews from a pilot study. Finally, the propositions were linked together using connectors based upon Schank's causal dependency theory (Schank 1972; Schank and Abelson 1977). Pennington (1981) notes that there was some ambiguity in coding links, because explicit natural language connectives between conceptually linked propositions do not always exist. In such cases, proximity in the text and coder judgment were used.

After the all coding was done, it was possible to organize the linked propositions into a causal event sequence, which could then be characterized in terms the episodic model presented in Figure 35. In this manner, story information was normalized in a manner that could be compared across participants, and combined to identify common structures across participants. This dataset was subsequently used as a basis variety of analyses, illustrating the centrality of story creation in juror decision making.

Pennington and Hastie (1986) found that there was a very high degree of consistency between stories constructed by jury members choosing the same verdict, and furthermore, that there were significant differences in stories between verdict categories. This was a powerful result, but the authors admitted the possibility that the method of data collection (an interview after a verdict was made) could have induced the story structure. Subsequent studies sought to illustrate that the relationship between story structure and verdict category was not just correlative, but causative.

In Pennington and Hastie (1988), it was argued that jurors spontaneously develop a mental representation that takes the form of a causal event model prior to the decision.

This argument was supported by demonstrating that jurors had better recollection for evidence items that were part of the causal model supporting their verdict. They also had higher false positive rates when queried about fabricated evidence items that were consistent with the presumed causal model. Another manipulation demonstrated that when evidence is presented in story order (temporal and causal ordering is preserved), jurors have an easier time constructing these mental representations and are more confident in their final verdict. Finally, Pennington and Hastie (1992) showed that the story construction process can be hindered by asking jurors to make a judgment after each piece of evidence is presented. In this case, juror judgments are better predicted by an “anchor-and-adjust” model (Hogarth and Einhorn 1992), whereby new evidence is averaged with the previous opinion. That is, each incoming piece of evidence was used to update the juror's opinion, rather than modify the juror's story.

The empirical data collected by Pennington & Hastie demonstrates with a fair degree of certainty that jurors indeed construct a causal temporal model of events obtained in testimony, and that this model is the basis of the verdict. As a final part of the EBDM model, four *certainty criteria* were proposed as a way of explaining how jurors chose between multiple stories. Certainty criteria govern the acceptability of a particular story, an individual’s confidence in the story, and their confidence in a final judgment. These criteria are:

1. Coverage, which describes how many pieces of evidence a story can account for. The greater the coverage, the more acceptable the story.
2. Uniqueness, which describes how many stories might fit the same pieces of evidence.

A unique story is much preferred.

3. Coherence, which consists of judgments regarding the consistency, completeness, and plausibility.
4. Goodness-of-fit, which describes how well the story fits the decision criteria.

While certainty criteria are a compelling addendum to the EBDM theory, there has not been a direct empirical validation of their accuracy.

I will adopt Pennington & Hastie's definition of a story, but future work will be required to build more concrete bridges between their work and mine. For the time being, Pennington & Hastie's work should be viewed as a backdrop to the work presented in the following sections.

2 A Narrative Analysis

Unlike in the previous chapter, the reader will benefit substantially from an understanding of the mystery for the following analysis. The full mystery and set of clues are attached as Appendix A, but the synopsis in Figure 36 offers a comprehensive summary; the solution to the mystery is included in Figure 37. The synopsis covers all of the clues, and the unshared clues are presented in colored text. Red clues were provided to EE, blue clues to BE, and green clues to ME.

Robert Blake has been murdered. He usually leaves for golf on Saturday around 6:30AM in the morning, but this Saturday, he was found dead, having apparently fallen from his porch, with a blunt force trauma wound to his head. His wife heard something around 6:40AM and then a car leaving. She looked outside, and saw Blake's car in the carport, but no one else's. She assumed the noise was Blake getting the truck out of the garage. She first found out about her husband when Eddie Stewart, the handyman came to the door at 10:30AM and told her to call an ambulance. She did this, went outside, and saw Blake's body.

Blake's wallet was missing, but later found at the QuickStop, without money or ID. The owner of the quick stop heard a "quiet car" peal away at 7AM, and went outside to find the wallet in the dumpster.

Mickey Marble is a business associate and old golfing buddy of Blake's. Lately, they've had a dispute about some bad auto parts Mickey has been supplying Blake. Blake has threatened to tell other people not to buy parts from Mickey. Mickey usually golfs with Blake on Saturday. This Saturday, he called Blake, and they argued. He drove over to straighten things out, but turned around before he got there. He got to the golf course at 6:50AM, a fact that is corroborated. While he was golfing, it occurred to him that maybe his new assistant, Louie Brown, had switched suppliers. He called him and left a message, a fact that was confirmed by Mr. Brown.

Billy Prince mows the lawn, usually on Saturday. His prints were found on Eddie Stewart's crowbar, which was found in the bushes, and his tire tracks on the driveway. He's known to have gambling problems, and to request frequent advances from the Blake's. He said he had just borrowed money on Friday (to explain his tire tracks on the driveway), and this was confirmed by Ms. Blake. However his tire tracks were determined to be from Saturday morning. When the detective confronted him about this, he came clean, and said that he had been there around 8AM to mow. Billy says that the crowbar was leaning against the door to the garage where the mower was kept, and that he had to move it out of the way to get the mower. When he saw the body, he "got outta there." Billy drives a car with a loud muffler.

Eddie Stewart is the handyman. His daughter Sue had worked for Blake for years, but recently quit after a suspicious argument. Eddie says he didn't ask about it. Eddie was at Blake's to tear down the barn at 6AM on Sat., but claims to have parked his car in the carport, gone directly to the barn, and to have left his crowbar out next to his truck. He also says he heard Billy's car around 7AM from about 200 yards away. According to Billy, Eddie typically is very careful about his tools. According to Ms. Blake, he is mostly deaf, and never wears his hearing aid.

Figure 36: Murder mystery synopsis

SOLUTION: Eddie did it because of something unspecified between Blake and Sue. He left his crowbar in front of the garage, waited for Billy to touch it, and threw it in the bushes, creating an appearance of means. He took the wallet, emptied it of money, drove to the QuickStop and tossed it in the dumpster, creating the appearance of motive. He claimed to hear Billy's car at 7AM, creating the appearance of opportunity.

Figure 37: Solution to the mystery

2.1 Focused Agreement

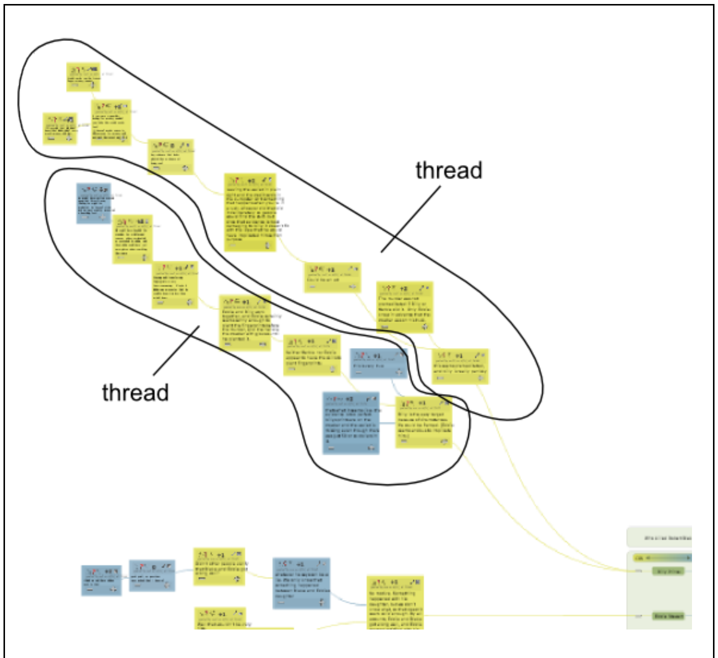


Figure 38: Threads in a non-adaptive group

Conversations in the non-adaptive condition have a different character than in the adaptive condition. Recall that within this domain, a “thread” refers to all descendants of a post that responds directly to a particular alternative (Figure 38). The conversations had by non-adaptive groups featured longer threads, more valence mistakes, and more “diplomacy” (people arguing both sides of the argument). An up-close examination of the conversations grounds these statistics in the observation that most of the non-adaptive groups seem to do all the heavy lifting of story construction in long, highly collaborative threads.

Most non-adaptive groups appeared to have at least one such thread in which most of the participants would get involved, connections between pieces of information would get made, and the backbone of the story laid down. A distinctive signature of these threads is that they contain a region of rapid posting, where the posts mostly support the same argumentative position (either by agreeing with one another, or disagreeing with a another post). In these areas of the conversation there is correspondingly little activity in other threads. I will refer to these as areas of focused agreement and the threads in which these areas are found as focus threads.

Post #	Time	Author	Valence	Post
45 of 93	14:10:01	ME	[top - Implicates M	I think it was either Mr. Marble or Mr. Stewart
57	14:13:45	BE	CON	so how do people stand on this? aka which one do you think is guilty?
60	14:15:20	ME	CON	I think Eddie is guilty but I'm still looking for a sufficient motive, probably with his daughter.
62	14:16:10	S1	PRO	I agree
63	14:16:33	S2	PRO	I agree with this, but I want some theory on how he got billy's prints on the crowbar
65	14:17:10	ME	CON [*wrong valence*]	Good point. I think it's possible that he removed it from the scene of the crime without thinking for some reason
66	14:18:21	BE	PRO	the crowbar was leaning on the garage door so billy had to move it to the side to get the lawnmower out, seeing how billy cam every saturday to mow lawn, eddie could have planned that
67	14:18:53	S1	PRO	BINGO
68	14:19:10	ME	PRO	I didn't know that
69	14:19:46	EE	PRO	AHA
70	14:20:24	S2	PRO	OK why not Eddie?
--thread continues--				

Table 23: Group 15_{na}'s focus thread; the focused agreement area is highlighted

Table 23 offers an example of a focused agreement area in one of the non-adaptive groups (Group 15_{na}). By this point in the conversation, B has been ruled out as suspect due to a lack of motive, and M alibi's has been discussed, but questioned. The inconsistency between E's claiming to have heard B's car and his hearing problem has been identified, and the possibility that E framed B has also been brought up, but has not gathered much support. Similarly, E's daughter being fired has been suggested as a possible motive, but only gathered partial support in the group.

This portion of the conversation is very important to the story formation process for the group, because it illustrates an instance where a piece of evidence or inference is found to lend support for a hypothetical story element that appears as part of a proposed story line. In line 60, ME suggests that E seems guilty, but that the current explanation of motive doesn't seem sufficient. S1 and S2 concur, and S2 adds that the absence of an explanation for B's fingerprints on the crowbar contributes to the lack of support for E's guilt. In response to this, ME suggests a *hypothetical explanation*, which would help preserve the theory that E committed the crime by explaining away the critical clue incriminating B. This leads BE to retrieve and explain the unshared clue that B had admitted to moving the crowbar. BE also weaves this clue into a possible story line for the rest of the group. All participants notice this, and acknowledge the importance of the information.

In the above dialog, the group may be seen to be “trying out” a story, by seeking to validate expectations that it creates. There are clear similarities between this process, and “expectation testing” portions of both Klein's RPD theory (Klein 1993) and Noble's situation assessment approach (Noble 1993) to decision making. When that expectation has been fulfilled through the application of data, the group coalesces around the story, and begins to try to test it for errors (post 70 “OK why not Eddie?”).

This observation is a “critical region” or area of story creation for the group. Once again, I will not be overly careful in defining what a “critical region” is. This is an avenue for future research. However, I will illustrate several of these regions throughout the following discussion. In each case, I will point out why they are more (or less) “critical” to the story creation process. Furthermore, as I will illustrate, areas of shared consensual agreement in the non-adaptive group frequently correspond to such critical areas of story creation.

From a purely quantitative standpoint, the shared consensual focus area presented in Table 23 has several features which make it unique within the context of the conversation as a whole. It is highly consensual – posts 62 through 70 contain no rebuttals (65 is posted with the wrong valence). All of the participants contribute at some point, which indicate that this is an important thread to the group as a whole. Moreover, posts 65 through 70 are sequential, indicating that there are no intervening posts elsewhere in the conversation – all activity is focused on this thread.

Note also that the posts appear quickly. The refresh cycle for the platform is five seconds, but, given processing time, it may take as long as ten seconds for a post to appear on other collaborators' displays. Given that there are roughly sixty animated nodes on the screen, and that the new posts that fly in from the side of the screen can take a little while to settle, any post interval on the order of a half a minute indicates that collaborators are actively awaiting the next move in the conversation. For the portion of the thread presented, then, the conversation has become effectively synchronous.

In the adaptive groups, similar instances of such threads are hard to identify. There are areas where critical pieces of the final group story get put together, but they are never characterized by rapid posting and highly consensual collaboration. Consider the thread created by Group 17_a, shown in Table 24.

Post #	Time	Author	Valence	Post
33 of 72	14:11:07	EE	[top level implicates E]	I think Eddie killed him heat of the moment over something with his daughter, and then tried to cover it up because he regretted it.
35	14:12:01	S2	CON	how could he have put billy's finger prints on his crowbar?
38	14:12:36	EE	PRO	Billy might have touched the crowbar when finding the body.
40	14:12:52	EE	PRO	Or Eddie could have paid Billy off to keep quiet.
48	14:17:25	ME	CON	That's not evidence.
54	14:18:48	BE	PRO	Yes. Billy admitted to touching the crowbar. That doesn't mean he used it to kill Blake.
55	14:19:56	S1	CON	Did he? I didn't have that.
59	14:21:51	S2	CON	why would he touch it?
62	14:22:59	BE	CON	Is there any evidence to support that idea?
64	14:23:41	EE	PRO	The missing money.
65	14:24:47	BE	CON	We don't know how much money was in Blake's wallet at the time of his death, only that he usually carried around about $50. Maybe he'd given all the cash in his wallet to Billy the night before.

Table 24: A critical story construction thread for Group 17_a

In some ways, the content of this thread parallels the one shown in Table 23. The unshared information about B touching the crowbar is hypothesized by EE (who does not have this information), and BE fills in the missing clue. But there are distinct differences in how this information gets used. Participants S1 and S2 both question the validity of the information[5], and even BE seems to miss the important connection between the crowbar and E's possible attempt to frame B. Contrast BE's utterance in post 54 of Table 24 with

5 It was emphasized to all groups that participants had different information.

BE's utterance in post 66 of Table 23. The BE user is in a position to make the inference that for the crowbar to be in B's way, someone (probably E) *placed* the crowbar there. This happens for Group 15_{na}, but not Group 17_{a}.

The thread for Group 17_{a} also takes much longer to evolve, is highly oppositional, and concurrent with other threads. Post 48 - “that's not evidence” - offers some insight. Groups in both conditions were instructed to concentrate on the evidence. However, in the adaptive groups it was explained that how evidence was represented in the system would determine the end result, and that if the participants represented evidence properly, the system would make the correct assessment. This may have resulted in a bias to evaluating evidence in isolation, rather than connecting the pieces. I will return to this in Chapter 8.

Conversations generated by the adaptive groups occasionally exhibit regions of rapid posting, where multiple collaborators are involved, and there is little activity in other threads. Such areas are indeed focus areas, but they do not correlate with story construction. The thread shown in Table 25 is once again from Group 17_{a}, and occurs roughly right in the middle of the thread shown in Table 24.

Post	Time	Author	Valence	Post
42 of 72	14:13:53	EE	[top level implicates M]	Do we have any more details about Marble that make him suspicious?
43	14:14:20	S2	PRO	yes
44	14:15:01	BE	PRO *(wrong valence)*	such as...?
45	14:15:15	ME	CON	and those details would be?
46	14:15:39	EE	PRO	Well, Blake was about to ruin his business.
49	14:17:32	BE	PRO	He was going to stop buying parts from Marble, but that wouldn't necessarily have ruined his business.
52	14:18:28	EE	PRO	My clues have him saying to the cop “he was about to ruin my business with other customers”
56	14:20:17	S2	PRO	and there was a letter that the victim had written to him stating that not only would he pull his business, but that he would expose his bad parts to other customers.

Table 25: Rapid posting region in Group 17_a; highlighted area is the area of rapid posts

The rapid burst of posts may indicate either excitement at the possibility of a new clue or perhaps frustration that the clue was not immediately forthcoming. The posts have little to do with the final story, though – they do not clearly exonerate M, nor do they contribute to the implication of E. The thread itself is an instance of a high degree of shared focus for the group, though, which is uncommon across the adaptive groups.

The above examples illustrate the subjective differences in story creation between the two groups. In the non-adaptive groups instances of shared consensual focus are common, pronounced, and frequently correspond with critical areas of story creation. Regions of shared focus occasionally occur within the adaptive groups, but they do not correlate with instances of story creation.

Because areas of shared consensual focus have a signature that is easy to identify, it is possible to use quantitative techniques identify them. In the following section, I describe such a technique, and present the results of its application.

2.2 Quantifying Focused Agreement

I have defined focused agreement areas in threaded dialog as those areas where:

- Several participants are involved in a single thread, and there is little activity in other threads.
- There is a high degree of consensus.
- Posts appears rapidly.

In the non-adaptive groups, inspection of the collected dialogs revealed that such areas are common across conversations and that they correlate well with critical areas of story creation among groups. In adaptive groups, regions that might be considered focus areas could not be easily identified, and when they were, they did not occur in threads that played a significant role in story creation.

Given the above definition, it is possible to create a metric based on the frequency and concentration of “agreeable” posts. The design of this metric flows immediately from the above definition. First, we may collect data in bins of equal temporal length along each thread. Focus is simply measured as those regions of the overall conversation where activity is high in a particular bin in one thread, and low in other threads. Agreement can be measured by counting the number of statements advocating each side of an argument in a particular bin and taking the difference between them. Areas of shared consensual focus are those areas where both focus and agreement are high. It is also possible to measure other quantities in this manner, such as areas of “spread agreement,” where conversation is highly agreeable but spread over many threads.

The general approach of binning data within conversations and analyzing frequencies of dialog moves was used to analyze group decisions by Poole (1981; 1983a; 1983b). Poole coded dialogs via a standard taxonomy of *interacts*[6], and binned data. He then used the binned data to identify phase transitions in group decision making activity – for example, from "orientation" to "evaluation." Poole used this technique to demonstrate that groups do not travel a single linear path when making decisions. This work contradicted the dominant viewpoint that groups follow a prototypical sequence of phases (Bales and Strodtbeck 1951), and was used as the basis of Poole's (1983b) multiple sequence model of group decision making.

The approach described here is related to Poole's, though somewhat less sophisticated. However, the underlying data, in its raw form, is richer because collaborators have organized their conversation into topic-oriented threads, and indicate within those threads which sides of an argument each utterance supports. Because of this, significant insights can be gained from the data with minimal processing.

To define these metrics more precisely, I will introduce several terms. First, let C be the total time of the conversation, δ be the size of each "bin" (to be determined by the analyst) and $B=C/\delta$ be the number of bins. As a convention, I will use b to identify the bin at index b (where $0 \leq b < B$). Let T be the total number of threads. Once again, as a convention, I will use the letter t to identify a thread at index t (where $0 \leq t < T$).

The conversation as a whole may be viewed as a $T \times B$ matrix, and we can now define

6 Adjacency pairs in conversation; Poole actually used and compared two different taxonomies, Bales (1950) Interaction Process Coding System and Fisher's (1970) Decision Proposal Coding System.

several measurements at each index in this matrix. First, let P_{tb} be the total number of posts appearing in thread t at bin b. P_{tb} is just a simple measure of activity. Also, let T_b represent the number of threads that show some activity in bin b. More precisely:

$$T_b = \sum_{\substack{0 \le t < T; \\ P_{tb} > 0}} 1$$

Let U_{tb} be the proportion of all possible users whom have made at least one post in thread t at bin b. The proportion of users contributing at a particular index in the conversation is a rough indicator of the amount of conversational activity (as opposed to just posting activity) that is occurring in a given thread at that point in time.

Finally, let A_{tb} represent the amount of agreement at each index in the conversation. To define this value, recall that each thread is anchored at a top level post that either supports or objects to a particular alternative. Let Y_{tb} be the number of posts appearing in thread t at bin b that agree with this top level post, and N_{tb} be the number of posts that disagree with the top level post. A_{tb} can be defined as:

$$A_{tb} = \begin{cases} \dfrac{|Y_{tb} - N_{tb}|}{P_{tb}} & \mathit{if}\, P_{tb} > 0 \\ 0 & \mathit{otherwise} \end{cases}$$

A_{tb} is thus a rough indicator of amount of consensus within a given thread at a given time. A value of 1 indicates that all posts are aligned at that index, and a value of 0 indicates that there are an equal number of arguments on either side of a given topic, or no posts at all.

We can now define composite metrics to characterize the type of activity going on in the conversation at each point in time, indexed by *b*. First, we can define focus as:

$$focus_b = \begin{cases} \dfrac{\sum_t U_{tb}}{T_b^{1+\varepsilon}} & if\, T_b > 0 \\ 0 & otherwise \end{cases}$$

The constant $\varepsilon : \varepsilon \geq 0$ is a penalty constant that allows us to measure the degree to which conversation is restricted to a single thread. When $\varepsilon=0$, $focus_b$ measures the average conversational activity over all active threads, and varies between [0..1]. However, when $\varepsilon>0$, $focus_b$ varies between $[0..1/T_b^{\varepsilon}]$, thus increasing our preference for activity that is isolated to fewer threads.

Focused agreement, or shared consensual focus, are areas of the conversation where both $focus_b$ is high, the interlocutors are expressing consensus opinions, and there is a lot of activity (recall, one of the features of focused agreement is rapid posting). Symbolically:

$$focused\_agreement_b = \begin{cases} \dfrac{\sum_t U_{tb} A_{tb} P_{tb}}{T_b^{1+\varepsilon}} & if\, T_b > 0 \\ 0 & otherwise \end{cases}$$

Once again, the constant ε plays the same role. Because $focused\_agreement_b$ includes a term for posting activity, it is no longer normalized, and varies between $[0..\sum_t P_{tb} / T_b^{1+\varepsilon}]$.

Note that when ε=0, the top end is just the average activity per active thread. It would be a simple matter to normalize *focused_agreement*$_b$ by letting P_{tb} represent the proportion of all posts in the conversation appearing at that index. The goal of such normalization would of course be to facilitate inter-group comparisons. However, normalizing over the total number of posts in a conversation introduces some semantics which run counter to the purpose of the metric. For instance, a group may be particularly "verbose," and post more frequently in parallel threads before and after the focus area than other groups. Normalizing would penalize this behavior (in comparison to other groups), even though the focused agreement area in such a case is more pronounced because it stands in sharper contrast to "normal" posting behavior. Developing a more general metric to facilitate inter-group comparisons will thus require further study and is an area for future development.

Finally, *spread_agreement*$_b$ is a general indicator of how much non-argumentative behavior is occurring in the thread that is **not** isolated to a single thread. This metric is similar to the inverse of focus, but does not consider the proportion of users that are contributing to the conversation, because it is not concerned with interaction between participants. Roughly, *spread_agreement*$_b$ might be considered to be characteristic of information seeking behavior.

$$spread\_agreement_b = \begin{cases} \left(\frac{T_b}{T}\right)^{\phi} \frac{\sum_t A_{tb} P_{tb}}{T_b} & if\, T_b > 0 \\ 0 & otherwise \end{cases}$$

The constant $\varphi : \varphi \geq 0$ performs a function analogous to the constant ε in *focused_agreement*$_b$. When φ *=0*, *spread_agreement*$_b$ simply measures the average agreeable posting activity (not necessarily conversational activity) occurring in active threads at bin *b*, and *spread_agreement*$_b$ varies from [0.. $\sum_t P_{tb} / T_b$]. As the value of φ is increased, *spread_agreement*$_b$ is penalized for activity that is restricted to a smaller number of threads, and varies from [0.. $\sum_t P_{tb} T_b^{\phi - 1} / T^{\phi}$].

In the following analysis, δ was set to two minutes, ε*=1*, and φ*=1*. These values were chosen by inspection. δ is based on the overall average rate of conversation, which is roughly one post every 35 seconds in either condition; the two minute bin size should, in general, capture three or four posts. Setting ε*=1* places a T_b^2 term in the denominators of both *focused_agreement*$_b$ and *focus*$_b$, heavily penalizing activity outside of a single thread. Finally, setting φ*=1* in *spread_agreement*$_b$ causes all agreeable posting activity in a bin to be to be averaged over the total number of threads.

For summarization purposes, results were averaged across threads, and then across groups in each condition to arrive at the results reported in Table 26. These statistics suggest that there is a persistent difference in focus, and particularly in focused agreement between the two populations. The differences in focused agreement cannot be accounted for by a general shift in "agreeableness," which if anything, seems to suggest that the non-adaptive groups are less agreeable.

	Adaptive	Non-adaptive	*p* (two-tailed T)
Average focus	.1	.13	*p<.01*
Average agreement	2.94	2.32	--
Average spread agreement	.13	.15	--
Average focused agreement	.09	.17	*p<.005*

Table 26: Summary statistics for computed metrics

Figure 39 (below) offers a closer examination of these three metrics for each conversation.

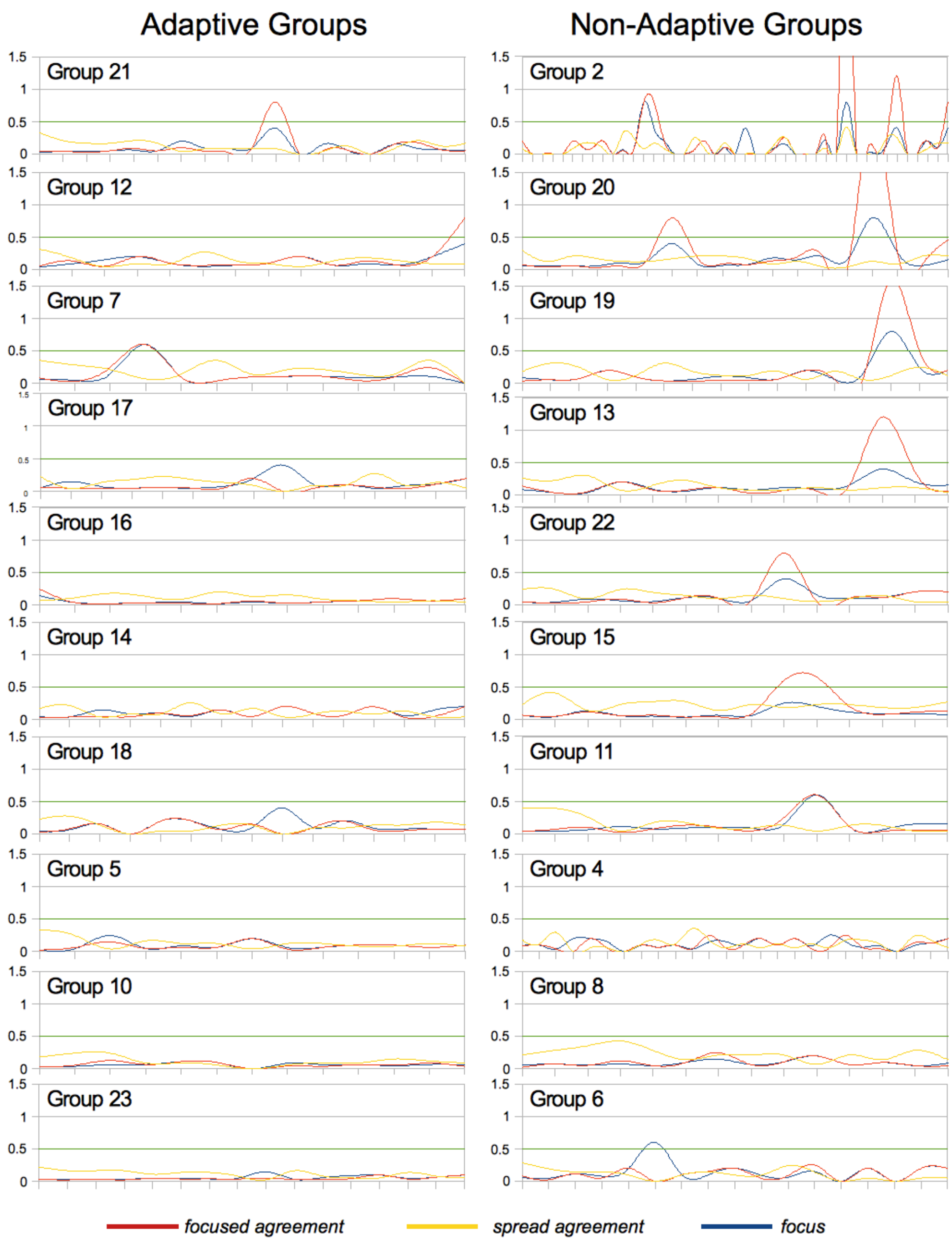


Figure 39: Focused agreement (red) in adaptive vs. non-adaptive groups; graphs in each condition are sorted top down by largest occurring peak. The horizontal at .5 is subjective cutoff for "interesting" peaks.

In the figure, y-axes have been set to a consistent scale (0-1.5), and lines have been smoothed to improve readability. The graphs in each condition are sorted top-down according to maximum peak size for focused agreement (red lines in the graph). Examining the data in each case reveals that most conversations in the non-adaptive condition have at least one strong peak of focused agreement, whereas most of the conversations in the adaptive case do not. For purposes of this analysis, I will define a "strong peak" as any peak in focused agreement exceeding .5.

Some general observations may be made about the data. Focused agreement areas in the non-adaptive groups tend to occur in the last half or even third of the conversation, which is suggestive of a "coming together" following prior, less focused activity. Accordingly, spread agreement tends to be higher earlier in the conversation indicating that many people are working in different threads and that there is little disagreement in those threads. This corroborates observations that the groups frequently started conversations by posting all their information in separate threads, and is roughly consistent with phase models of group decision making (e.g. Poole 1983b).

Focused agreement and focus are generally well correlated (as would be expected), but this is not always the case. Groups 17_a, 5_a, and 6_{na} all exhibit peaks in focus that are not correlated with peaks in focused agreement. These are areas of focused *disagreement.*

Finally, note that the individual cases discussed both in this chapter and in the previous chapter appear to be properly represented. Group 17_a exhibits a focus area that begins at 14:13:08 and ends at 14:17:08, but does not correlate with focused agreement, as discussed above. Group 15_{na} exhibits a broad area of focused agreement from 14:13:16 through 14:19:16, corresponding to the focused agreement area presented in Table 24. Furthermore, recall that neither group 6_{na} nor 8_{na} (groups with the lowest peaks of focused agreement on the non-adaptive side) were able to complete the mystery. Finally, Group 23_{a} (the group with the lowest peak in focused agreement on the adaptive side, and the lowest average focused agreement overall) was an adaptive group in which two participants exhibited highly contentious posting behavior.

These initial observations make intuitive sense, but an analysis of the conversation is necessary to determine if indeed focused agreement correlates with collaborative story formation. Such an analysis is presented in the following section.

2.2.1 Focused Agreement and Story Creation

The analytical approach to quantifying focused agreement identifies peaks of activity in most of the non-adaptive groups. For Group 15_{na}, the region identified by this approach has previously been shown to correspond to a critical area of story creation (Table 24). Groups 6_{na} & 8_{na} are consistent with the pattern, as the absence of a focused agreement peak reflects the inability of a group to arrive at a compelling collaborative story. All but two of the remaining non-adaptive groups are compatible with the basic pattern. The two groups that do not fit the pattern are Group 4_{na}, which does not show a distinct peak, but does construct a robust story, and Group 22_{na}, which does have an identifiable peak, but does not seem to use this area to construct a solution to the problem.

In the following sections, I describe just one positive case in the non-adaptive groups to clarify the pattern (Group 2_{na}). However, additional examples are provided in Appendix

B for the interested reader. Following the positive example, I will examine each of the adaptive cases to illustrate that the identified peaks do *not* correspond to areas of story creation. I then turn to the negative cases on the non-adaptive side and perform a slightly deeper analysis.

2.2.1.1 The Positive Prototype

Group 2_{na} is unique in couple of ways. The story that is created by the group is very sparse (see Figure 40 for a synopsis). There is in fact very little that can be explicitly identified as a story in the conversation, and the group seems not to do a very effective job weaving the clues into a narrative. However, the group has some of the strongest peaks of focused agreement, and each of these are critical to the final solution in some fashion. The first peak (not included here) corresponds to a discussion of B's gambling problem and the stolen wallet, and rejects the notion that he stole the wallet on the basis of the small amount of money that would have been taken and the fact that he had just gotten money the night before.

> Eddie did it; he lied about the muffler, and tried to place the blame on Billy. There was that thing with his daughter – that's unclear, but it's a possible motive. Plus, he left his crowbar out when he usually locks his tools up. There are other inconsistencies which don't seem to make a lot of sense about him. Billy is a liar, but it's unclear he stole the wallet (why not take the credit cards?). Still not sure why his fingerprints were on the crowbar. Mickey really didn't seem like the type to do it – had an old friendship with Blake, and could've probably patched things up.
>
> *Figure 40: A synopsis of Group 2_{na} s story.*

The remaining two peaks are the critical components to the group's final story, and immediately precede the final vote. By the point these peaks occur the group has dismissed M as unlikely because he had previously been friends with the victim, and could have probably taken care of the dispute regarding the bad auto parts. The group is a little less clear about B. His dishonesty is problematic for them, as are his fingerprints on the murder weapon. However, there's some question about his motive, and the group seems to think he would have stolen the credit cards from the wallet to feed his gambling addiction. With respect to E, they don't seem to think his daughter's quitting is a very good motive, however they do think it's suspicious that he left his crowbar out, especially given that he was usually careful with his tools. Several minutes into the conversation S1 (who does not have the clue about E's poor hearing) mentioned that it seems unlikely E could have heard a muffler all the way from the barn, but no one replied to that. The area of shared focused agreement occurs 24 minutes after that observation. The conversation that corresponds to both of these peaks is shown in Table 27.

Post	Time	Author	Valence	Post
38 of 53	19:23:52	EE	[top level implicates E]	What's this thing about Eddie having a hearing problem (in the 2nd interview w/ Mrs. Blake)?? Anyone have any thoughts on that or was that just randomly thrown in??
39	19:24:44	BE	PRO	Eddie could be lying about hearing the muffler
40	19:24:46	S1	PRO	explain more. copy and paste that part cuz i dont have that evidence
41	19:25:18	S1	PRO	probably, if he had a hearing problem how could he hear such a muffler from far away?
42	19:25:23	ME	PRO	I don't have that part in my interview, but that is very interesting
The following post appears in a different thread at this point.				
45	19:33:22	ME	[implicates E]	1st interview Sgt. C.:We're trying to get some things about last Saturday sorted out. You said you got to the Blake's about 6 in the morning and went straight to the barn. Then about 7 you heard a car? with a loud muffler. Mrs. Blake thought you came to the patio door around 10:30. Is that about the time that you discovered Mr. Blake's body? Ed. S.:I'm not sure about that. It could have been around then. I really don't remember.
The following continues the original thread.				
46	19:33:59	EE	PRO	Lt. M: One other matter? Is it true that Eddie Stewart has a hearing problem? Ms. B: Yes, he is very hard of hearing. Sometimes when he gets a phone call, I have to call him. I've tried calling to him from the deck, but he never hears me. I have to walk right up to him before I can get his attention.
47	19:34:32	EE	PRO	Lt. M: Doesn't he have a hearing aid? Ms. B: He has one, but he doesn't wear it while he is working. He says that it doesn't fit well. It's one of those tiny ones and he's afraid he will lose
48	19:35:00	S1	PRO	then how could he possbily hear the muffler from the barn?
49	19:40:59	ME	PRO	I think this is important, he didn't have his hearing aid with him but claims to hear Billy's loud muffler

Table 27: Focus thread for Group 2_{na}; focused agreement areas are highlighted

In this instance, the focus thread is split, as EE goes back to the evidence to copy and paste into a posting. During the gap, there are a couple additional postings made in another thread by S1 and S2 regarding the timing of events as described by the various characters in the mystery. At 19:33:52 (nearly eight minutes after the "gap" began), ME posts another relevant piece of information, but in a different thread. Immediately following this, EE posts the relevant information. S1 and ME both take note of this, and voting occurs soon after.

As with Group 15_{na}, Group 2_{na} follows a similar pattern of establishing an expectation - "Eddie could be lying about the muffler" - and then filling this expectation in with data, in the form of text copied and pasted from the mystery affidavits themselves. Unlike with Group 15_{na}, the expectation is not quite so obviously a story element, and occurs in response to the uncertain evidence posted by EE in line 38. Nonetheless, this phase of the conversation serves a very important role for the group, and they make their final decision soon after this exchange takes place. Hence, this area seems to be a critical story creation point for Group 2_{na}.

2.2.1.2 The Adaptive Groups

The focus graphs shown in Figure 39 illustrate that several of the adaptive groups do have identifiable peaks. Group 17_a has been discussed above, and it was found that the focus area was an extended debate about a piece of evidence that contradicted the dominant story. The remaining groups are examined below.

Group 21_a

Group 21_a "gamed the system," and chose M as the guilty suspect, contradicting the system's base assessment of B. Group 21_a also has the most prominent peak of all adaptive groups. The peak occurs between 16:07:37 and 16:09:37 (about twelve minutes into the conversation) and the only activity occurring at that point in the conversation occurs in the identified thread. The entire thread is included in (Table 28).

Post #	Time	Author	Valence	Post
51 of 86	16:07:06	ME	[top – exonerates M]	He called Brown at noon on Saturday asking they'd switched suppliers. Why would he care if he had killed Blake that morning?
52	16:08:00	EE	CON	deniability?
53	16:09:04	S1	PRO	Or b/c he wanted to know if it would sound plausible that Blake had agreed to go back to him.
54	16:11:47	S1	[top – implicates M]	The place where the wallet was found is in the Eastwood Mall, which is on the way to the golf course....

Table 28: Focused agreement thread for Group 21_a

The thread (posts 51-53) functions to explain away one of the clues exonerating M (M called his parts supplier after the golf match), and so supports the final story. However, rather than serving as a critical element of story *creation,* this exchange is discounting evidence that might otherwise conflict with the dominant story. In this instance, the identified peak is not especially meaningful.

Group 12_a

Group 12_a was one of the adaptive groups that put together the correct solution, but did so in part because the group was heavily biased towards E going into the discussion. The focus area begins at the very end of the conversation, in the last thread (Table 29).

Post #	Time	Author	Valence	Post
43 fo 47	16:26:32	ME	[top – implicates B]	my evidence says that billy denies ever picking up the crowbar
44	16:26:56	BE	CON	my evidence has him later admitting to it
45	16:27:01	EE	CON	my evidence says that his fingerprints were found on the crowbar.
46	16:27:21	S1	PRO	admitting to what?
47	16:27:48	BE	PRO	Moving the crowbar

Table 29: Focused agreement thread for Group 12$_a$; focus area is highlighted

Only posts 45 and 46 are included in the focus area, because the preceding posts overlap occupy a bin with additional activity. However, the thread occurs rapidly, and by the proposed definition of shared consensual focus, posts 43 through 45 should also be included in the focused agreement area. Nonetheless, the function of the thread is to argue away incorrect (or rather incomplete – ME is missing the unshared information that B picked up the crowbar) evidence that would seem to contradict the dominant story. This is relevant to the story, and could be a critical piece of evidence, however, it occurs at the very end of the discussion, after the group has apparently reached a consensus, and voting has begun. Thus, this exchange has little actual bearing on the story construction process.

Group 7$_a$

Group 7$_a$ chooses B, and does not develop a very sophisticated analysis of the mystery. The focused agreement thread identified is shown in Table 30. As with the previous examples, the identified thread discusses a possible piece of evidence. In the thread BE is seeking further support for B as the suspect. The other collaborators agree that B is the suspect, but disagree with the logic of BE's particular line of reasoning. The thread does not contribute in a significant way to collaborative story creation.

Post #	Author	Valence	Post
3	BE	[top – implicates B]	Blake died from the fall not the crowbar. so whoever killed him was either not strong or not using full strength. Billy was sick that morning, so he could not have hit blake hard enough to kill him.
8	EE	PRO	I disagree. He was not sick, he later admitted he was actually there doing work.
14	BE	CON	he never said he was lying about feeling ill.
15	EE	CON	if he was well enough to do manual labor, he could have hit someone with a crowbar.
16	ME	CON	that's irrelevant what matters is that he lied about being there.
17	BE	CON	it does matter because it would help explain why the victim died the way he did
-- thread continues --			

Table 30: Focus thread for Group 7_a; focus area is highlighted

These examples, and Group 17_a above, illustrate that focus areas in the adaptive groups do not have much do with story creation. That is not to say that story creation does not occur, or that these focus areas are not indicative of other collaborative phenomena. I will return to this possibility in a later chapter, but for now will turn my attention to the non-adaptive groups.

2.2.1.3 The Exceptions

Both Group 22_{na} and Group 4_{na} get the correct answer, though Group 4_{na} collaborates more effectively. However, Group 4_{na} does not appear to have a substantial peak, and the peak for Group 22_{na} has little to do with the final story.

Deeper examination reveals that these groups have a couple of unique features. Posts were longer, containing more complicated and complete utterances than were found in the other groups. In Group 22_{na}, this seemed to reduce the degree of collaboration; the conversation was dominated by two individuals with longer posts (S2 and EE), and the final result arose from the efforts of these individuals. Table 31 presents a portion of the dialog that is representative of the general conversational tone in Group 22_{na}. The frustration apparent in BE's post (post 29) stems from having posted the same information several times, but never having received an acknowledgement.

Post #	Author	Valence	Post
23	S2	[top – implicates B]	Needed money for gambling issues. Blake's wallet was stolen. I can't see Marble or Stewart needing money. However, it's entirely possible that Stewart commited the deed and then Prince looted the body before taking off. This is what I believe so far.
28	EE	PRO	I agree. What I'm wondering about now is the cars- lets pool our information on what cars everyone says were there when. we know billy's car was there and left around 8 (it has a loud muffler)
the following post appears in another thread at this point			
29	BE	CON	YO!!! His fingerprints were on the crowbar! What can explain this????? He said he touched the crowbar..ummmmm sketchy

Table 31: A representative exchange for Group 22_{na}

Metrics supporting these subjective observations about the two groups are shown in Table 32. Note that while the variance for inter-post interval appears high in the table, this is mostly due to Group 2_{na}'s cutting and pasting behavior, which unnaturally inflated their inter-post interval (to an average 94.67 seconds), and skewed results for the non-adaptive condition. The average inter-post interval without Group 2_{na}'s data is also included in the table for comparison.

Metric	Group 4	Group 22	Non-adaptive average (sd)
Avg Post Length (chars)	93.07	119.31	66.67 (21.76)
Follow-on Ratio	.25	.1	.23 (.11)
Inter-post Interval (sec)	39.83	31.06	32.64 (22.99) / 25.75 (7.78)

Table 32: Various metrics comparing Groups 4 and 22. Follow-on ratio is the ratio of arguments that follow those in the same thread to all arguments in a conversation. Inter-post interval is the average

amount of time between posts. The second number in the inter-post interval average column is without Group 2 data.

To summarize, both groups had much longer than average posts, but Group 4_{na} had a correspondingly high inter-post interval, and normal follow-on ratio. Group 22_{na} had a slightly higher than normal (omitting Group 2_{na} data) inter-post interval, but a very low follow-on ratio. Based on this analysis, we can hypothesize that the lack of identifiable focus areas in Group 4_{na} might be attributable to a slower posting rate in general, and could be rectified by increasing the bin size of the technique. Conversely, the problem with Group 22_{na} probably cannot be rectified in this manner, and it is likely that the peak is an artifact. As the following analysis illustrates, these hypotheses are borne out.

The two graphs for Group 4_{na} are shown in Figure 41. The change in bin size does not appear to change the magnitude of the peaks significantly and the peaks appear to line up with similar peaks at the smaller bin size. However, the increased bin size smoothes the data, and three relative peaks become apparent.

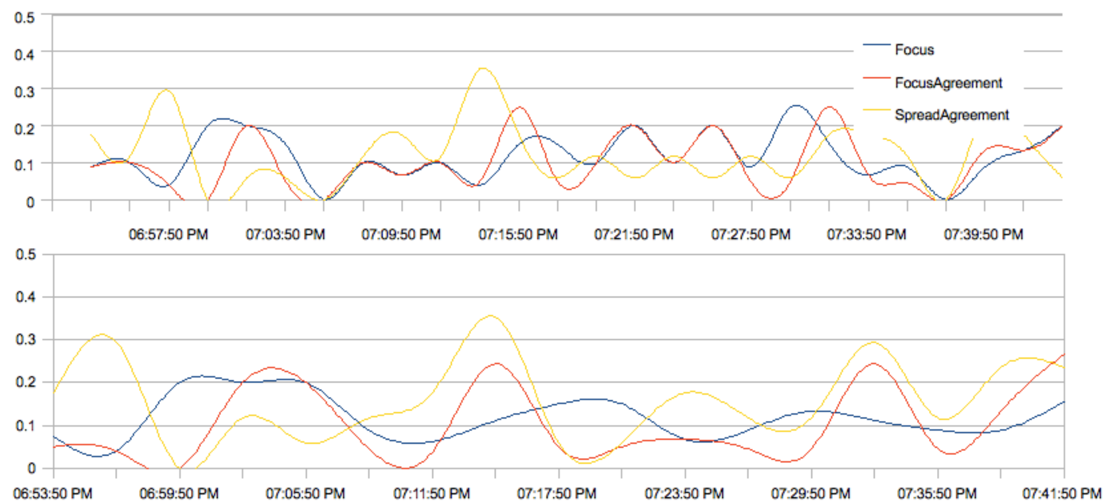


Figure 41: Different bin sizes for Group 4. (a) Top - δ=2 minutes. (b) Bottom - δ=3 minutes.

A synopsis of the story constructed by Group 4 is shown in Figure 42. The group uncovers some of the unshared clues about Eddie, but not about Billy. Given the dearth of available information, the group manages to put together a reasonably accurate narrative.

> Eddie did it – he definitely lied about hearing Billy at 7, and there's a good chance he set Billy up. There was also all that time during which he didn't tell anyone about the body – what was he doing? And he left his crowbar behind, even when he was supposed to be tearing down the barn. Not sure what his motive was (something with the daughter? something about Billy using his tools?) but this is enough to implicate him. Billy did have money / gambling problems, lied and all that. And his fingerprints were on the crowbar. But he's nervous and panicky, and doesn't seem bright or evil enough to commit murder like this. Mickey kind of had a motive, but he has an alibi – though it's not clear that this alibi is air tight.
>
> *Figure 42: Story synopsis for Group 4_{na}*

Increasing bin size has several effects. Note that focused agreement is no longer quite as well correlated with focus, but that it is better correlated with spread agreement. This is because the "agreement" term in the current equation, which is a simple sum across threads, begins to outweigh the focus term. However, the average metrics over all bins have not changed significantly from the case with smaller bin sizes. In particular, average focus (over all bins) does not change at all.

Upon examination, it appears that each of the areas identified by the peaks are important areas within the conversation. Conversation corresponding to each of these areas are shown in Tables 33, 34, and 35. Note that some posts that lie within the the focused area bin have been omitted because they occur in different threads and are unrelated to the dominant conversation.

Post #	Time	Author	Valence	Post
16 of 72	19:03:38	EE	[implicates E]	Seems very eager to put all the attention he can on Billy. Further, Billy is an easy target: twitchy, a gambler. Is Eddie deflecting attention from himself?
The following post occurs in a different thread				
17	19:05:03	S1	[exonerates B]	The murder was not premeditated if Billy or Marble did it. Only Eddie knew in advance that the crowbar was on his truck.
The following continues the original thread(s).				
18	19:05:35	BE	CON	No motive. No alibi either, though.
19	19:05:36	ME	PRO	and he knew exactly what to say in the first interview, but couldn't respond to follow-up questions. it fits with the premeditation theory.
--thread continues--				

Table 33: The first focus agreement peak for Group 4_{na}; suggests support for E having framed B.

In Table 33, we see the beginnings of a story taking form with EE's suggestion that E may be trying to deflect attention from himself. This is of course not a crime, but suggests that maybe E is trying to cover one up. It serves an important role in that brings the “framing” story into the group's conversation. Posts 17 and 19 lend support to EE's proposed story. Neither of these posts are actual “evidence,” but they illustrate that other members have taken it upon themselves to try to find support for the framing story.

Post #	Time	Author	Valence	Post
29 of 72	19:14:47	BE	[implicates E]	maybe he wanted to avenge his daughter
30	19:15:16	EE	[top - Implicates E	Eddie has a hearing problem, according to Ms. B, and he doesn't wear his hearing aids while working. Does he claim at any point that he heard something important, that was difficult to hear? This could be a sticking point in his story.
32	19:15:35	BE	PRO	who quit working for Blake because of some issue between them
33	19:16:38	S2	PRO	that's a good point. the hearing aid thing is an interesting development as well.
*The following post occurs in a different **sub**-thread (an earlier part of the original).*				
34	19:17:23	BE	[implicates E]	motive – to avenge his daughter. Maybe Blake did something to her.
The following continues the original thread(s).				
35	19:17:26	S1	PRO	YES! he IDs billy's car by its sound
				--thread continues--

Table 34: The second focus agreement peak for Group 4_{na}; suggests motive for E and identifies critical connection between hearing problem and car muffler; post 33 is omitted

The post in Table 34 is more obviously a critical area in the story construction process. As with Group 15_{na} (the first example in the chapter), EE declares an expectation that would fit with the current dominant story, based on a piece of unshared information (E's reluctance to wear his hearing aid). In line 32, BE registers this as an important element, and S1ultimately fills in the expectation with a piece of data. Note that while this exchange is occurring, BE is building support for E's motive.

Post #	Time	Author	Valence	Post
53	19:31:21	EE	[exonerates E]	Didn't other people verify that Blake and Eddie got along well?
The following post occurs in a different thread				
54	19:32:22	BE	[implicates E]	he also says he went back to get his crow bar. that puts him there. It's not like he doesn't have gloves if he's doing work in a barn.
The following posts continue both threads as indicated				
55	19:33:03	BE	PRO [*wrong valence*]	good point, no questions were asked that
56	19:33:50	S2	PRO	I think we only have eddies word to that.
57	19:15:35	BE	PRO	He says his job was to tear down the barn, which would certainly require a crowbar. However, that is never either confirmed or denied by anyone else, as far as I know.
*The following post occurs in a different **sub**-thread (parallel to 53)*				
59	19:35:16	S2	Implicates E	That certainly fits. Eddie is possessive about his tools: maybe he didn't want other people working with them, or something similar.
--thread continues--				

Table 35: Third focused agreement peak for Group 4_{na}; many points covered, supports suspicions about the missing crowbar, and offers one of the two possible motives suggested for E. Post 58 is omitted because it occurs in a different thread and is unrelated to the topic matter in the other posts.

The area identified in Table 35 is perhaps less critical, and less cohesive than the other regions described above. Several things occur, however, that are related to construction of the final story. Lines 53, 55, and 56 serve to eliminate "evidence" that detracts from the dominant story – specifically, EE's impression that Blake and E got along well. Line 57 is performing a similar function, by explaining away a potential (though incorrect) problem with the dominant story. Finally, Line 59 is responding to an alternative motive that is proposed outside of the area shown in the table.

Thus, in the case of Group 4_{na}, increasing the bin size helps to sort relevant story conversation from the other conversation. The technique correctly identifies several important elements, however it also captures substantially more activity in other threads that is unrelated. Thus, in broadening the bin size, we capture relatively more agreement, and relatively less focus. This is discussed further below.

Manipulating the bin size for Group 22_{na}, unlike with Group 4_{na} above, effectively eliminates any peaks from the focused agreement graph. The results of this manipulation are shown in Figure 43, and the reason the peak is eliminated is because the focus term has been greatly reduced. Whereas in the case of Group 4_{na}, average focus over all bins did not change, in Group 22_{na} average focus is roughly half what it was. The difference is not quite significant at the .05 level ($p=.057$), but is substantial enough to flatten the curve. Considering the overall low follow-on ratio for Group 22_{na}, and the collaborative dynamic that appears to be in effect in the conversation, this result is not unexpected, and confirms the hypothesis that expanding the bin size would not capture more relevant story construction.

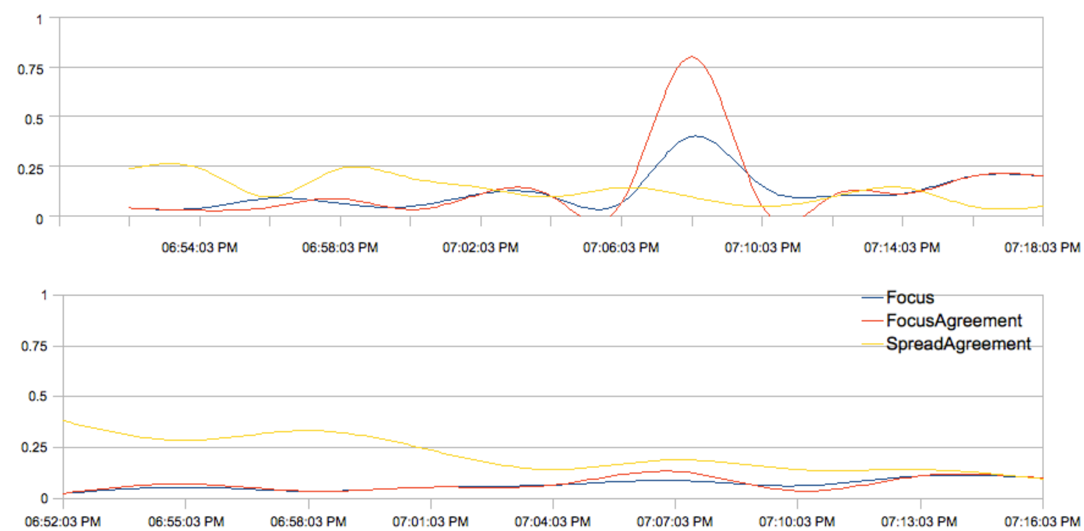


Figure 43: Focused agreement for different bin sizes Group 22_{na}; (a) Top - δ=2 minutes.; (b) Bottom - 3 minutes.

In conclusion, "focused agreement" is a useful and important phenomenon in group decision making for the domain under investigation. In addition, the quantitative approach supported by the proposed definition appears to accurately identify critical areas of collaborative story creation, when decision making is mediated by an argument network representation and decision makers are engaged in semi-synchronous collaboration.

The approach raises many questions and points to avenues for future research. With regards to the quantitative metric, it is clear that it is subject to artifacts and errors. One type of artifact is due to a static binning approach, which draws boundaries at regular intervals along the conversation. As a result, portions of a focused agreement area can get split, under-weighting potentially interesting areas. This type of sampling error might be rectified with more sophisticated binning strategies. A related problem is that groups with different base posting rates, as with Group 4_{na}, might require different underlying bin sizes.

This latter observation leads directly to questions regarding the consistency and boundaries of the focused agreement phenomenon across groups with different collaborative dynamics. By broadening the bin sizes for the Group 4_{na} data, we "cast the net" a bit wider, at the cost of precision. The resultant relative peaks did correspond to instances of story creation, but they also captured relatively more activity in other threads, as well as activity in parallel sub-threads of the same parent thread. More investigation is required to determine if this indicates a revision to the original definition of "focused agreement" is necessary, if the results obtained in Group 4_{na} should be discounted, or if the quantitative approach simply needs refinement.

Finally, the measurement of focused agreement allows us to reflect back upon the differences between adaptive and non-adaptive groups, and helps to illustrate that decision-making groups without a mediated decision process generally engage in a different kind of story creation than groups subject to such mediation. In the following sections, I will examine the effects of the different dynamics upon the narrative product that results from a group's collaboration.

2.3 Collaborative Story Creation

As described at the outset of this chapter, story creation occurs in both conditions. This observation is well supported by Pennington & Hastie's work on juror decision making. As I have illustrated in the preceding analysis, critical elements of story creation occur in highly collaborative regions of the conversation for the non-adaptive groups. However, this does not appear to be the case for the adaptive groups.

To complete this analysis, I will address several outstanding questions. First, how do critical areas of story creation function in terms of the overall decision making process? Furthermore, if story creation occurs in both conditions, but critical pieces of story creation do not occur in highly collaborative focus areas for the adaptive groups, where do they occur? Answering these questions will then support a final discussion regarding the overall effect of the mediating platform, and the results obtained in the preceding chapter.

2.3.1 Story Elements

The summary provided in Figure 35 and Figure 36 at the outset of the chapter describe both the clues provided to the collaborators, as well as the story as it “really” happened. The story itself ascribes motivations to the main character (covert revenge for some implied wrongdoing to E's daughter), and describes an unfolding of events that illustrates how these motivations become actions and events that have effects in the world and leave behind evidence. The process the collaborators go through to solve the mystery involves reasoning “backwards” from evidence of these events – both observed, and reported – to the events themselves. The glue that binds these events together is a story.

Within the collected conversations, the process by which a group uses evidence to choose (or construct) a story is characterized by a flow of information through three levels. Consider the three statements in Table 36, all taken from Group 13_{na}, at different points in the conversation.

Post #	Time	Author	Valence	Post
34 of 84	19:09:41	BE	[top - implicates E]	His daughter worked for Blake – she quit, things ended poorly.
56	19:16:13	S2	[implicates E]	this is big motive especially because she was “very upset” who knows what went on between sue and bob?
59	19:16:36	S2	[top - implicates E]	revenge killing?

Table 36: Evolution of evidence in conversation

Post 34 is the first statement that offers E's daughter's argument with Blake as evidence, and it is offered as something that casts suspicion upon E. It may imply much more, but the utterance, by itself, is simply a statement of evidence. Post 56 follows this same evidence, via an inference that there may have been some unspecified event between E's daughter and her boss that led E to become angry, to support the idea that E had a motive for murder. Finally, post 59 sees the attribution of motive become an explicit element of a story – that is, that the murder was a consequence of a revenge killing.

The “revenge killing” story may have existed in one or more of the collaborators' minds long before the utterance in post 59, and was perhaps implied by both posts 34 and 56. In practice, though, the movement of information though the various levels suggested by this analysis – evidence, through inference to motive or event, to story – can be witnessed in the phrasing of information as explicitly stated by the interlocutors. Hence, a conservative interpretation is warranted, and implications can be ignored.

The process by which a given story becomes more or less dominant among multiple possible stories flows with the movement of information from evidence to plausible story. Pieces of evidence motivate inferences, which in turn modify support for events or motivation. Possible events and motivations are not a story until they are placed within an episodic context, such as the one described by Pennington and Hastie (1986).

This process can be observed and described by identifying the relevant pieces of evidence, inferences, and story elements in conversation, and mapping their movement through the conversation. I have performed such an analysis for one particular story line, for a selected set of groups to illustrate the effects of the conversational differences described in the previous section. This analysis is presented below. Note that though I choose to focus on one story in order to highlight information flow within conversation, this does not mean that only one story is under consideration by the group. A more exhaustive analysis is a potentially rich avenue for future work.

2.3.2 An Approach to Story Based Analysis

The story I have examined is the “correct” story, as presented in Figure 36. The goal of this analysis is to illustrate how information flows through a group to become the story that ultimately determines the group decision. To perform this analysis, we will need to identify the critical bits of information in the conversation, track them as they flow through the levels described above, and identify allegiances in the group with respect to this information. Thus, our first steps are to identify the relevant pieces of evidence and inferences to be tracked, and then to develop a more rigorous definition of a story element.

The evidence to be tracked is drawn primarily from the clues that are described in Appendix A. Several additional pieces of information are included because they were found to be used to support inferences in the collected conversations. Inferences, for the “sleuthing” domain, are of three types:

- Inferences from inconsistencies that reduce support for hypothetical events or motivations, and corresponding inferences about the motivation of actors responsible for such inconsistencies.
- Inferences, from evidence, that increase support for motivations or events.
- Inferences, from evidence, that reduce support for motivations or events.

A description of each piece of evidence and inference for the “correct” story is shown in Table 37.

Story			
E did it. It was a revenge killing, because of some implied indiscretions between RG and E's daughter. E may have planned it. He attacked RG with crowbar, and then proceeded to set B up; mentioning hearing B's muffler at about 7am to to detective, planting the crowbar deliberately to get B's fingerprints on it, and taking money out of the wallet and ditching the wallet in the dumpster.			
Motive	**Implicates E**	**Exonerates B**	**Exonerates M**
JIE1 - Something to do wih his daughter – quit under suspicious circumstances **UIE1** - argument with Gates **JIE1+UIE1 →** **Em** – E had anger because of implied wrongdoing to his daughter	**SIB3 -** E says heard B's muffler **UIE2 -** E has hearing problem **SIB3** + **UIE2** → **EI1i** - framing B by implying he had opportunity (muffler)	**JEM1** - M arrived at GC by 7 (weak) **UEM1** – M arrived at GC before 7 (stronger) **UEM2 -** wallet dropped off at 7 **JEM1/UEM1+UEM2 →** **ME1i** - M could not have dropped off wallet (no opportunity to frame B)	**UEB3** - Car dropped wallet off was quiet **UEB3 →** **BE1i** - B could not have dropped off wallet
	JIE2 - E's time unaccounted for **UIE3 -** Car not seen at house **BE1i + ME1i + JIE2 + UIE2 →** **EI2i** - E framed B by creating the appearance of motive (wallet)	**JEM2** - M left at 6:30 **UEM1** - arrived before 7 **JEM2+UEM1 →** **ME3i -** alibi checks out (weak)	**JEB2** - RG always gives him money **UEB1** - MG confirmed giving him money on Fri **JEB3** - wallet contained only 50$ **JEB2+UEB1+JEB3 →** **BE3i** - B has no motive
	SIE4 - Left crowbar out **UEB2 -** Billy touched **JEB4** - B didn't throw **SIE6** - E usually locks up tools **SIE4+UEB2+JEB3+SIE6 →** **EI3i** - threw crowbar in bushes, framing Billy, giving him means.	**UEM3** - M called LB after GC **UEM3 →** **ME2i** - M was interested in resolving the problem (no motive)	**JEB1** - B (claims) to have showed up at 8 **JEB1 →** **BE2i** - B did not have opportunity

Table 37: Evidence and inferences to be tracked in the construction / selection of the correct story for the murder mystery

Acronyms for each element in the table are constructed as follows. If an acronym represents a piece of evidence, the first letter is either "S" for shared, "U" for unshared, or "J" for shared information that was not part of the set of clues described in Appendix A. The second letter indicates whether the evidence is considered to be implicating (I) or exonerating (E) evidence, the third letter indicates the suspect the evidence is considered to be about (E, M, or B). The fourth character is a unique identifying digit. If an acronym represents an inference, the first three characters are derived in the same manner as the last three for pieces of evidence, followed by an "i" to indicate that it is an inference. If an acronym is a positive inference to motive for the crime (there is only one), it is constructed from a letter indicating the suspect that is implicated, and the letter "m."

The inferences to be tracked are those that occurred in practice. They are described in the table as **A + B → C**, where A and B are either evidence or prior inferences, and C is the resultant inference. For my purposes, it has not been critical to classify inferences according to a typology, as was done in Pennington and Hastie (1993). The application of a framework, such as Collins and Michalski's (1989) logic of plausible inferences, may be productive and would allow this analysis technique to be applied to more generally. However, in this investigation, inferences may be identified in collected transcripts using the provided dictionary.

Story elements are not included the coding dictionary embodied in Table 37, because there are too many possible story elements to enumerate. Roughly speaking, story elements are those instances of dialog that indicate that a story is beginning to take form. More precisely, I define story elements as those that can be identified by cues in the dialog that:

- Invoke a theme that identifies a prototypical story – such as "revenge killing" or "framing."
- Establish an ordering of events in time.

- Extend either of the above via the introduction of hypothetical or actual evidence.
- Summarize any of the above elements.

My definition of a story element is slightly different from Pennington's (1981), primarily because I include the invocation of "prototypical" stories, and summarization. Pennington's coding scheme may be able to handle such elements through other means, but they were not explicitly identified in Pennington (1981).

Because story elements may extend or summarize others, it is necessary to evaluate utterances within context to identify them accurately. Thus, by itself

"EE: Billy might have touched the crowbar when finding the body."

is a hypothetical piece of evidence of type UEB2. However, within the context of the following exchange:

EE: I think Eddie killed him heat of the moment over something with his daughter, then tried to cover it up because he regretted it.
S2: how could he have put billy's finger prints on his crowbar?
EE: Billy might have touched the crowbar when finding the body.

it becomes apparent that this statement is an introduction of hypothetical evidence designed to support a story line that has been proposed.

Once a conversation has been tagged in this manner, allegiances may be tracked by noting rebuttals or statements of support that are attached to the tagged items. Where no such replies occur, allegiances cannot be assumed. Furthermore, I will assume that individuals do not change their minds unless there is explicit evidence of this (e.g. an utterance indicating a shift in position).

2.3.3 Two Analyses

The above coding scheme was used to annotate two complete conversations, one from the non-adaptive Group 13_{na}, and one from the adaptive Group 17_{na}. This coding scheme has not yet been validated, and is the subject of future work.

Type	Utterance	Code(s)
E	claimed to hear billy's car, even though he's clearly deaf	SIB3, UIE2
E	He handled it to get to the mower - claims he never used it, though.	UEB2
I	he was at the golf course before the others got there at 7 and the wallet was dropped off at 7	UEM1, UEM2,ME1i
E	His daughter worked for Blake - she quit, things ended poorly.	JIE1
S	eddie said he heard the loud muffler, and said it was billie's. was he trying to frame him?	SIB3,UIE2,EI1i
E	Only thing that'd make sense is his daughter's termination at Blake's business	JIE1
I	this is big motive especially because she was "very upset" who knows what went on between sue and bob?	JIE1, E1m
I	if billy's only motive was money; he didn't seem to get very much	JEB3, BE3i
S	revenge killing?	JIE1, UIE1
I	Also, Blake was *very* nice to Billy, always giving him advance payments	JEB2, BE3i
E	Billy said he moved the crowbar to get to the mower.	UEB2
S.ext	no, but eddie might of used the opportunity to frame him	UEB2, EI3i
I	He never said he threw it into the bushes, just that he moved it - Eddie could move it	JEB4, EI3i
S.ext	Eddie also wants to pin this on Billy from the beginning. (muffler)	EI1i
S.sum	possibly angry father + lack of emotion about murder + easy scapegoat opportunity	JIE1,Em, SIB3, UIE2, EI1i, UEB2, JEB4, SIE6

Table 38: Tagged dialog for Group 13. "E" is evidence, "I" is inference, and "S" is story. "S.ext" extends another story element, and "S.sum" summarizes several story elements.

Table 38 presents each of the tagged elements from Group 13_{na}'s conversation. Items are listed in the order that they appeared in conversation. Codes in the "Type" column are either "E" for evidence, "I" for inference, or "S" for story. "S" may be appended with ".ext" or ".sum" to indicate extension, or summary, respectively. Note that evidence, inferences, and story elements are arrayed roughly in that order over the sequence of utterances captured.

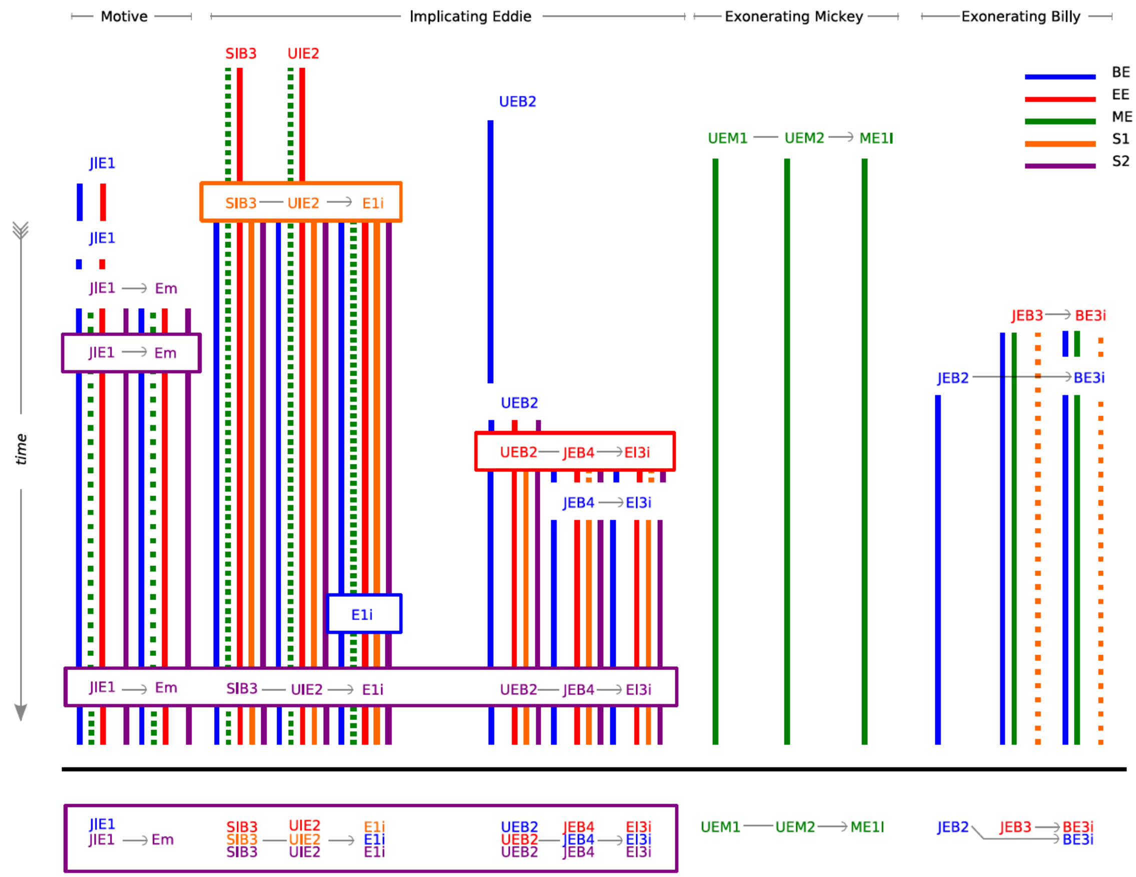


Figure 44: *Visualization of story creation for Group 13$_{na}$*

These elements, along with the allegiances of each individual, are captured in the visualization shown in Figure 44. The visualization may be interpreted as follows. Time flows from top to bottom, although only order is represented faithfully. Each participant is assigned a color, as indicated in the key. The color of each code represents the participant that introduced that element to the conversation. Bars flowing from each code represent our knowledge of the different participants' allegiances, based on replies to posts containing tracked elements. Solid bars indicate support, dashed bars indicate opposition, and the absence of a bar indicates no assessment could be made, either because no response was posted or the response was ambivalent (e.g. a question). Some approximations have been necessary, as there is some lag time between the introduction of an element and the expression of allegiance by another participant, and time is not represented in the visualization. Hence bars are visually connected to the elements they are about, even though a supporting or rebutting statement may in fact appear several minutes later.

Arrows between elements indicate that an explicit inference was made. Note that while it does not occur in the present cases, it is possible for a single utterance to contain both an information element, and an inference, but not connect the two, so the arrows are necessary. Parenthetical elements indicate the existence of that element was hypothesized, such as in “Billy may have touched the crowbar...” Finally, boxed elements are story elements. Note that only items in the “motive” and “implicating Eddie” categories can be parts of the final story. Other elements rebut other possible stories (e.g. either M or B did it).

Across the bottom of the visualization is a representation of the final story, as constructed via this analysis. This final representation, which I will refer to as a “story genome” that contains the entire lineage of a story, and inferences that contradict other stories. Stacking order indicates the order in which each item was mentioned (the top element in a stack is was mentioned first), and the color of each item indicates which individual that was responsible for the item. The color of the box is determined by the *last* individual to describe the story.

Reflecting back upon the analysis presented above, the area identified by the focused agreement metric is exactly the third story creation point shown in the visualization (the red box). It is clear in Figure 44 that this story creation point has no explicit detractors, and three of the other members express explicit support for the story element, confirming that this is indeed an area of focused agreement.

Post	Time	Author	Valence	Post
70	19:20:30	S2	[top level implicates E]	very much disregarded the question did he find the crowbar
73	19:21:34	S1	CON	how would billie's fingerprints get on eddie's crowbar?
75	19:22:30	BE	CON	Billy said he moved the crowbar to get to the mower.
76	19:23:06	S1	CON	and would that really require throwing the murderweapo- crowbar into bushes, to be hidden?
77	19:23:51	EE	CON	no, but eddie might of used the opportunity to frame him
78	19:24:10	BE	CON	He never said he threw it into the bushes, just that he moved it - Eddit could move it
79	19:24:48	BE	PRO	Eddie also wants to pin this on Billy from the beginning. (muffler)
80	19:25:16	S1	CON	what's eddie's motivation to frame billy?
82	19:26:04	BE	CON	It's somebody that is not Eddie.
83	19:2632	S1	PRO	lolpwnd
84	19:27:06	S2	CON	easy scapegoat

Table 39: Focus thread for Group 13_{na}.

This region of the conversation is shown in Table 39. Although the highlighted area contains only three posts, the binning process has prevented the metric from grouping posts 75, 76, and 80 into the focus area. All of these posts are taken together illustrate a critical region similar to the one illustrated for Group 15_{na}, earlier in the chapter. The bit of unshared evidence has previously been mentioned, but not connected to the “framing” story. Posts 77 through 79 are precisely where this connection takes place.

The final story for Group 13_{na} is truly a collaborative result. Different participants have "touched" different information elements on their journey from evidence to story, and no single individual can be said to have "ownership" of the final product. There is not, however, unanimous support for the final story. ME is a persistent detractor of both Em and EI3i (E's motivation and E's framing of B by claiming he heard his car). However, ME is also responsible for the inference that reduces support for M, and supports one of the inferences that reduce support for B. Notably, ME's final reported post-discussion preference is for E.

The results of the same analysis for Group 17_a are very different. As with Group 13_{na}, all coded information elements for Group 17_a are presented in Table 40 and Figure 45.

Type	Utterance	Code(s)
E	claimed to hear billy's car, even though he's clearly deaf. His comment about hearing Billy's car is suspicious, considering his hearing issues.	SIB3, UIE2
E	The argument between Eddie's daughter and Blake is sort of suspicious.	UIE1
I	But there was probably only $50, and he didn't take the credit cards.	JEB3, BE3i
E	Also, the store owner who found Blake's wallet heard a car outside his shop, but only because the tires screeched. It wasn't because of the muffler.	UEB3
E	Billy said Eddie was a stickler about his tools, but Eddie said that he had left his crowbar next to his truck. Doesn't sound like a stickler to me...	SIE4, SIE6
I	But Blake had given him money the night before.	UEB1, BE3i
S	I think Eddie killed him heat of the moment over something with his daughter, then tried to coerer it up because he regretted it.	UE1, Em, SIB3, UIE2, EI1i
S.ext	Billy might have touched the crowbar when finding the body.	UEB2
S.ext	Mrs. B heard a car leaving in the morning. It may have been Eddie trying to get rid of the evidence.	EI2i
E	Also, it was after 6:40, and she thought her husband should be gone by then	EI2i
S.ext	Yes, Billy aditted to touching the crowbar. That doesn't mean he used it to kill Blake.	UEB2

Type	Utterance	Code(s)
E	And he really makes a point about bringing up Billy, and mentioning his crowbar being missing.	SIE4, UEB2

Table 40: Tagged dialog for Group 17_a

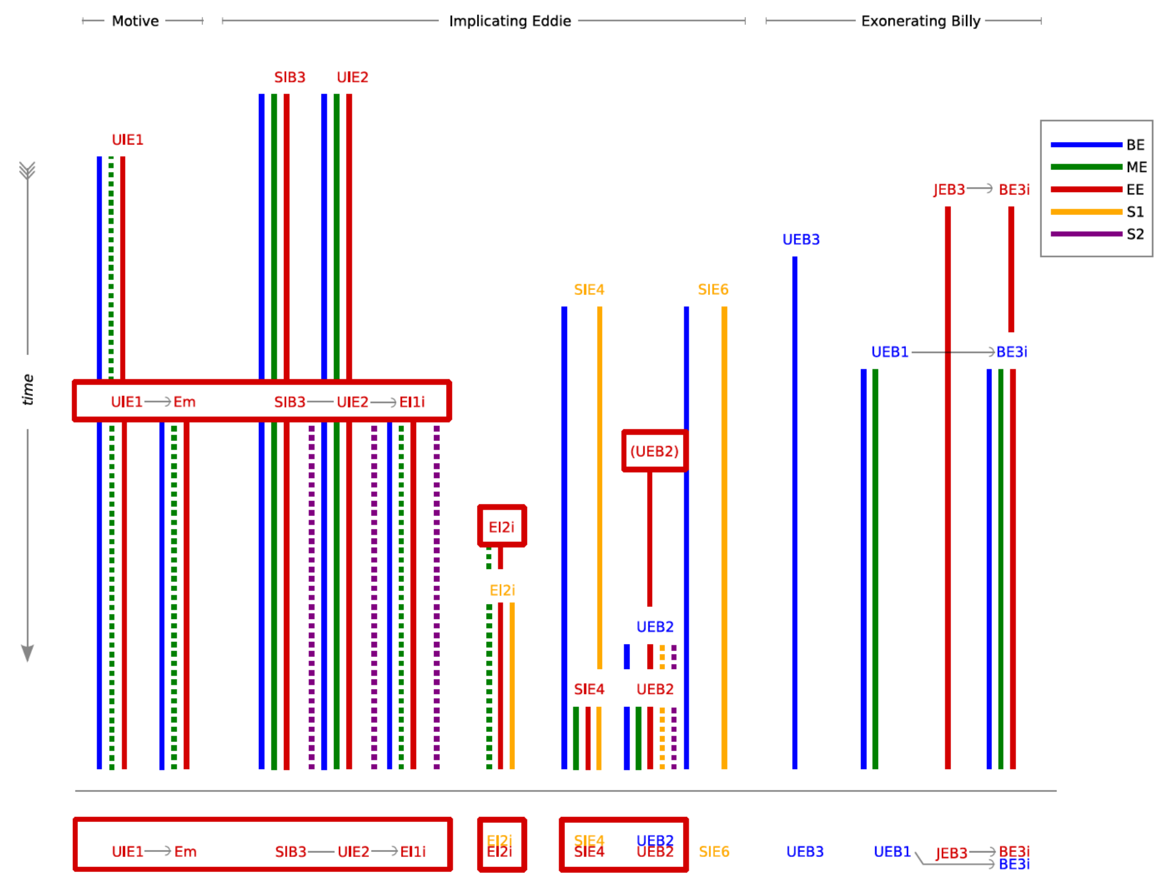


Figure 45: Visualization of story creation for Group 17_a

It is apparent that the final story produced by Group 17_a is substantially less collaborative than that produced by Group 13_{na}. EE is responsible for articulating most portions of the story. The group's final decision is also less consensual than Group 13_{na}'s; only one other participant supports E's motive and the inference that E framed B by creating the appearance of opportunity, and a different participant agrees with the (ungrounded) inference that E had something to do with the wallet. The inference that E framed B by leaving the crowbar in his way is never completed, even though the unshared information about B touching the crowbar is discovered. Thus, we can see that Group 17_a manages to form a story because one individual is able to put together enough information to do so.

The process of generating similar analyses for the other stories has begun, and initial findings suggest that those illustrated here are representative. In the following and final discussion of this case study, I will draw the covered material together and reflect back upon the results presented in previous chapter.

3 Discussion

The preceding analyses offer further confirmation that adaptive mediation fundamentally alters the group decision making process, and in a significant manner, and we can now say with more precision what these differences are.

Groups of decision makers with an adaptively mediated decision process (the adaptive groups) tend not to be very collaborative. They do not have highly focused, consensual discussions. Stories play a role in in the decision process, but the platform itself did not mediate story construction. Instead, stories are built primarily by individuals who are able to put the pieces together, and argue their case via the provided mediating platform. This is not to say that the platform did not have an effect on how information was evaluated by individuals in the construction of stories, nor should it imply that story construction either always follows or always leads the evaluation of evidence. However, it does mean that "putting two and two together" is an act that occurs more often within individuals rather than among members of a group using the adaptive platform.

This is not necessarily a negative. This aspect of the platform seems to encourage a more argumentative norm, and such norms have been shown to improve information processing from the perspective of the common knowledge problem (Postmes, Spears, and Cihangir 2001). Furthermore, it demands that whatever story wins has the balance of supporting information. As long as people can identify the right weights and valences for the "bits" of information – as generally, it seems they can – the mediation will guarantee that the best supported option – according to the bits of information, anyway – wins.

The non-adaptive groups, on the other hand, were highly collaborative. Stories were constructed through cycles of focused agreement, where collaborators came together to weave their individual pieces of information together to create a truly joint product. Because there was no need to defend this product from a mediated decision process that can only understand separate, weighted pieces of information (and not stories) there was no need to carefully ground the storyline in bits of supporting data. Instead, the group was able to focus on connecting information in conversation. As a result, the pieces of information exchanged in conversation, when isolated and counted, did not seem to support the final outcome.

This is not necessarily positive. There were certainly instances where groups had an excellent collaborative process, but became enamored of a bad story, which attracted incorrect bits of supporting evidence, and led to the wrong answer. And, although the numbers were small, lack of a good collaborative story seemed to have a more damaging effect on the non-adaptive groups than the adaptive groups.

In the balance, it seems that mediating the decision process improves accountability to the data, but underutilizes or even impairs collaborative information processing. It places a damper on the kind of conversational connectedness that allows insights to occur between people, rather than within people, and hence reduces the availability of this process as a resource for decision making. For some sorts of decision problems, this may be exactly what we want to do, and for others, precisely the wrong thing.

In summary, then, I have developed a platform based on a rationalistic model of decision analysis, designed to make the collaborative decision making process more like an ideal in order to overcome a problem that has been identified in the literature. The platform was quite successful in doing just this. However, it may be that the problem examined exhibits features which require a different sort of support. What this implies for future attempts at such mediation is discussed further in the following chapter.

Chapter 7: Discussion & Future Work

This dissertation has illustrated how it is possible to leverage a groupware platform's privileged position as a mediator in a collaborative task both to solve a knowledge acquisition problem and to guide a collaborative process towards an ideal. This was done by illustrating the application of these techniques in two separate case studies. Both of these successfully illustrated the described techniques, and this work can serve as a resource for other developers who wish to employ them. In the course of demonstrating these contributions, I have developed a novel approach to group decision support, and uncovered some exciting and unexpected findings that are suggestive of a rich avenue of investigation into group decision making.

In the following section, I will revisit my main contributions, and in so doing point out constraints that apply and gaps that need to be addressed, as well as some promising avenues for further research. Following this discussion, I will revisit the topic of group decision support specifically. Finally, I will conclude with a brief observation about methodology in the design of adaptive groupware.

1 Adaptive Groupware

The work presented in this dissertation has addressed two questions facing the designer of adaptive groupware. First, how do we obtain the runtime information required to drive the adaptive functionality? Second, what should the system do with that information once it's been obtained? Two suggestions are offered here: first, to design or identify a mediating representation that captures the necessary information; second, to use that information to manipulate the users' collaborative process in order to improve it with respect to some ideal. There are some constraints in implementing these strategies, and there are also some opportunities to extend these approaches. I will discuss both of these in the following sections.

1.1 Knowledge Acquisition

The knowledge acquisition approach introduced in Chapter 3 is based observations and insights offered by Alterman (2000) about the development of adaptive (or autonomous) systems. First, as with any mediating artifact, an adaptive system is part of the triad consisting of the designer, the user, and the system itself. This has been called the *mediation triangle* (Cole and Y. Engeström 1993). To consider the relationship between the system and user as a relationship between observer and observed is an arbitrary distinction that ignores the user's role in the mediation triangle. It also makes things a lot harder for both the system (at runtime) and the designer (at design time), because the user plays no role in providing the system with the information it needs to perform its function.

Acknowledging the role of user in the mediation triangle addresses the knowledge acquisition problem, but does not offer any insights regarding *how* the designer might solve it. A second observation made by Alterman (2000) is that the partnership between system and user that exists at runtime is an asymmetric one (see Terveen 1995), and to consider the communication that goes on between the system and the user as a “conversation” is to require the user to bear the brunt of the communication costs (see Suchman 1987). If instead we consider the the system to be a part of the distributed cognition of the joint human-computer system (Hutchins 1995b; D. A. Norman 1991), our attention is drawn to the fact that human and computer have different properties with regards to information processing. Rather than attempt to design a “dialect” that will allow the system and user to converse about the task, the designer might focus upon allocating information management responsibilities between the user and the system in a manner that makes the best use of their respective capabilities. In so doing, the designer solves part of the knowledge acquisition problem (by distributing some of the user's runtime task information into the system), while improving the task from the user's point of view (by reducing the amount of work he has to do).

This is what the incorporation of a CR accomplishes. However, it would be misleading to suggest that the development of a CR is an approach to solving the knowledge acquisition problem for the *purposes* of implementing an adaptive strategy. If no CR exists for a given domain, the development methodology discussed in Chapter 3 advocates studying the use of a platform that is equipped with generic communications tools in order to identify what kinds of information can and should be distributed into the system, and several analytical methods have been developed to support this endeavor (Alterman et al.

2001; Feinman and Alterman 2003; Feinman 2006). However, the knowledge acquisition "needs" of a pre-slated adaptive strategy have no bearing upon this analysis.

Hence, the development of CRs for a given groupware domain is not part of a methodology for developing adaptive systems. It can, however, drastically simplify the subsequent deployment of adaptive technology. A very useful avenue for future work would be the identification and classification of CRs that are specific to different domains of collaborative activity. Such a catalog would be a boon for interface designers in general, as well as for those who wish to develop adaptive systems with more purposeful intentions.

1.2 Adaptive Mediation in Groupware

The development of adaptive functionality in any system implies that the designer has some theory or ideal vision of what a user's domain activity should look like when it is mediated by the designed tool. This is often the "software interaction ideal," which suggests that a piece of software should allow its user(s) to effect their intentions in the task domain as directly as possible. As has been noted by other's, (e.g. Bodker 1989) this ideal is equivalent to Heidegger's notion of "ready-to-hand" (Heidegger 1962). This ideal is the tool-builder's ideal, and its attainment may or may not be possible via the introduction of adaptive support.

However, this is a limiting view of the potential uses of adaptive technology. Adaptive systems can be used to extend human ability and improve domain performance beyond what would otherwise be possible. There are a handful of classes of adaptive software that do this, and each may be considered to be representative of an ideal. Safety critical systems and even more mundane error correcting systems exemplify an ideal of perfect action. Information retrieval systems exemplify an ideal of omniscience. Intelligent tutorial systems exemplify an ideal of perfect knowledge transfer or perhaps effortless learning.

In groupware, the field remains relatively unexplored, perhaps because the ideals that might be considered with respect to collaborative activity are harder to specify with the precision that is required to build adaptive strategies. Research in education and learning is a rich source of information that may be used in this regard. The work with PIERCE/Epsilon mentioned in previous chapters rests upon this tradition. Goodman et al. (2005) built upon the work of Soller (2001) and Soller and Lesgold (1999), whose work in turn built upon a much larger body of research that examined the correlation between speech acts and and effective knowledge exchange (e.g. Baker and Lund 1997).

However, as illustrated with Epsilon, attempting to align a knowledge acquisition strategy with an analytical method developed elsewhere is not always straightforward. The approach presented in this dissertation did not leverage a theory of an ideal collaborative *process*, as may be found in the group processing literature (e.g. Hirokawa and Poole 1986), for precisely this reason. Rather, I began with a theory about an ideal decision *outcome*, which I could determine at runtime given the mediating structure offered by argument visualization, and used this model to influence the collaborative process. The approach has led to some new insights about group decision making, and I report on these in the following section.

2 Group Decision Support

The REASON platform is a novel approach to decision support that overcomes some of the problems with modern systems. As discussed, some possible reasons why existing systems have not been well-received has been that they are simply “too complex,” and provide little concrete payoff.

I have suggested that the problem of complexity is in fact the knowledge acquisition problem in a different guise, and have addressed it in REASON in a couple of ways. Instead of serving as an oracle to which users make offerings of data so that it can produce a correct answer, the system serves as a mediator that shapes the decision making process to make it more like a rational ideal. The interface structure is compatible with a relatively natural deliberative process, and allows people to speak through, rather than to, the system. This aspect of REASON is a characteristic of its role as a coordinating representation in the deliberative process.

The other way REASON addresses what may be perceived as complexity in other GDSSs is that it does not produce a final answer, but instead ties the collaborators' final answer to a deliberative process. The system is a partner and traffic cop, but the final answer is something that the users choose. This aspect of REASON is a characteristic of adaptive mediation.

I have also offered a careful analysis of how REASON changes the decision making process. These empirical results indicate that if a decision problem can be effectively broken down in a standard MCDA fashion, and collaborators are able to deliberate for an extended period of time in synchronous manner, REASON will guide collaborators to a decision that reflects their collective evaluation of the evidence. As I will discuss below, not all problems may be broken down in such a manner and not all instances of deliberation occur in synchronous, non-time-critical settings. However, the findings presented here are a concrete result that can be offered to others as a justification for why the system may (or may not) be appropriate for a given decision task. Within the field of research on GDSS systems such a result is a significant contribution.

The above contributions are substantial, but the approach illustrated by REASON and subsequent analysis of its use are just first steps along the path to a much deeper investigation. There are many areas which deserve further study, or may present opportunities to extend the techniques I've employed. I will discuss some of these in the following sections.

2.1 Algorithm

The underlying algorithmic technique used to drive the the decision platform can and should be formally validated in several ways. Informal validation studies indicated that, for a variety of argument network topologies, the system's assessment mechanism produced the correct answer. However, different topologies have different semantics, and this will impact the way the assessment occurs. Consider the topology in Figure 46:

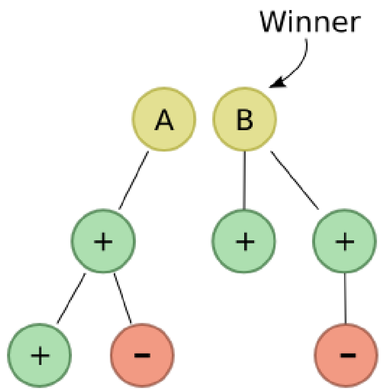


Figure 46: A possible network topology

In the figure, there are the same number of arguments for and against each of the two alternatives, yet “B” is the winner, because the “con” argument beneath A explains away a larger proportion of the aggregate evidence in favor of “A.” This is not necessarily incorrect, however it does imply a subtle semantics, and further analysis will be required to characterize these semantics. It may be possible to do this inductively (e.g. by extrapolating from the aggregation algorithm itself), but this might also be done empirically by sampling the space of topological possibilities.

Along similar lines, it would be valuable to determine error tolerances within a group of collaborators. That is, how many incorrect assessments can the system handle, and still achieve the correct answer? This question broaches a much larger topic regarding the system's capacity to overcome bias in groups. The system makes it difficult for any one individual to overwhelm the system. If we assume relatively even posting volume across members of a group, what proportion of a group would it take to overwhelm the system and push an agenda through? What if we did **not** assume even posting volume? These are very interesting questions, and require further study.

The algorithm itself may also be extended. Researchers have criticized Dempster's aggregation rule, because it can behave oddly under conditions of extreme disagreement, leading to aggregate results that none of the participants believe in (see Zadeh 1984, for an example). Although such cases are unlikely to occur with many participants, several alternative aggregation strategies have been developed to address the problem. An evaluation of these alternative methods in the context of the decision support platform developed here would be a worthwhile investigation.

2.2 Analysis

In addition to validating the algorithm itself, several of the techniques used to analyze the collected transcripts would benefit from further scrutiny and validation. First, the initial coding of the transcripts to identify clue-relevant information and wrong valence posts will need to be verified by independent coders before these results can be considered to be concrete. However, there was little subjectivity involved in these assessments, and it is unlikely that any significant differences will be found.

With respect to the story coding analysis used in the last part of Chapter 7, efforts should be made to establish a standardized set of coding guidelines, as the categories applied did require some subjective judgements to be made. The quantitative metric for shared consensual agreement did offer some validation that the items coded as story elements in conversation were indeed significant events, because these events were identified by hand prior to the development of the metric, and the metric was able to identify them with a high degree of accuracy.

Nonetheless, complete validation will require being able to communicate the coding scheme to others. As discussed in Chapter 7, there are many similarities between the coding scheme I employed and Pennington's (1981) approach. I offer some additional categories, but Pennington's “explicit story reference” and “evaluative / hypothetical” may be more clearly defined than some of my categories. A profitable future endeavor might be to explore a combination of the two approaches.

With a more complete coding framework (and access to willing coders), the entire corpus of collected data might be coded in order to extract multiple story lines. Such an effort would enable closer examination of the negotiation that goes on in a group trying to select between multiple possible stories. Ultimately, this might offer the basis for a set of theoretical criteria that govern the selection of a story by a group of decision makers, similar to that proposed by Pennington and Hastie (1992). It might also lead to the development of new mediation structures, as described below.

2.3 Focused Agreement

The metric developed for focused agreement presents many exciting opportunities for future investigation. A more thorough coding of the collected transcripts, as discussed above, would enable a more complete analysis of the relationship between focus, agreement, and collaborative process. Establishing a correlation between simple quantitative metrics, such as speed of posting and thread focus, with phases in the evolution of a collaborative decision making process would have enormous value, both for the evaluation of teamwork, and for future development of mediating artifacts. For instance, what would be the effect of providing a team of decision makers with continual ambient awareness to let them know when something “critical” was happening? Such a feedback mechanism is related to Carroll et al.'s (2003) notion of “activity awareness,” and holds potential as another type of adaptive mediation for guiding group processes.

It will be necessary to establish the boundaries within which the "shared agreement" analysis is viable. At the very least, it requires interface structure that allows people to separate different topical contributions into different "physical" areas of a conversation. Some of the work presented here has also indicated that the base speed of posting within a group makes a large difference in the evaluation of focused agreement. More sophisticated algorithms, which can detect and account for such variability may be able to help with this. Along these lines, it would be particularly interesting to see whether or not such an analysis can be applied to less focused asynchronous collaboration. Some of the work analyzing Wikipedia change logs (e.g. Viégas, Wattenberg, and Dave 2004) suggests that such an analysis might be possible.

2.4 Different Decision Problems

In Chapters 6 and 7, it was shown that the type of adaptive mediation used appears to improve accountability to the data, but underutilizes or even impairs collaborative information processing. It places a damper on the kind of conversational connectedness that allows insights to occur between people, and hence reduces the availability of this process as a resource for decision-making.

In the case of "murder-mystery" solving, and very likely in cases of jury decision making, this can be problematic. It may not always be, though, because collaborative decision making does not always involve collaborative story creation. In fact, the majority of hidden profile experiments study those those kind of tasks that do not require inferences to be made between information items. Wittenbaum et al. (2004) offer the following survey of types of tasks used in hidden profile experiments that examine the common information problem:

1. The best student body president candidate.
2. The optimal job candidate to hire.
3. The best company for investment.
4. The best drug to market.
5. The correct diagnosis of a medical case.
6. The guilty suspect in a murder mystery.

The first three problem types make up the bulk of the work on the common information problem, and the first four are decision problems where the choice can be made in standard MCDA fashion – that is, the criteria do not need to be connected to make a decision. The fifth case, which was described in Larson et al. (1996) is potentially more interesting. The decision materials were derived from the QMR (Quick Medical Reference) expert system, which provides ranked diagnoses for given sets of symptoms. Unfortunately, this study offered no analysis that might reveal if symptoms could be examined in isolation, or if certain tuples of symptoms had special significance in the diagnosis process.

The rationalistic model that was used to mediate the group's decision may be more appropriate in the other types of cases typically examined in hidden profile studies, precisely because information elements in these studies do not need to be combined to form inferences, and the rational model emphasizes decomposition. Fraidin (2004), examined the murder mystery employed here, and demonstrated that dyads were better able to make better decisions when individual members had pieces of information that needed to be connected to form an inference. However, within the literature on the common knowledge phenomenon, there as been very little discussion of this aspect of decision problems.

Outside of that literature, and primarily within the literature on naturalistic decision making, others have commented on the tendency of traditional decision analysis approaches to focus on simplistic problems that do not occur in realistic settings. Hammond (1993) not only offers some particularly harsh criticism, but also quantifies what he considers to be the important kinds of uncertainty in naturalistic domains, through application of the *Lens Model* (Hammond 1988). The model focuses upon the uncertainties between observer and cue, cue and criterion, and uncertain potential relationships between the cues themselves. Hammond places decision problems on a continuum, and suggests that the location of of decision problem on this continuum determines the degree to which *analytical* or *intuitive* cognitive processes should be used. These can be seen as analogous to the central and peripheral processes identified in the group information processing literature.

In my data, two processes were used by groups in either condition. Rather than characterize one as rational and the other as intuitive, it might be more accurate to say that one was was decompositional, and one was integrative. Furthermore, both of these processes were important. This lead to the possibility of a hybrid approach to decision support for the class of problem represented by the one studied here. Such an approach is discussed in the following section.

2.5 A hybrid approach

While it is probably not a reasonable near-term goal to support all kinds of decision-making within a single platform, the work presented in my analysis suggests that some synergies may exist between narrative / integrative style decision making and rationalistic / decompositional approaches. In my case study, I found that groups with adaptive mediation did in fact make decisions that were consistent with the information the exchanged, at the expense of a richly collaborative story construction process. The non-adaptive groups, on the other hand, had very rich collaborative engagements but did not seem to evaluate all of their data. However, both conditions got the correct answer about the same amount of time. Finding the right way to support both of these aspects simultaneously may allow groups of decision makers do better than those in either of the two conditions I studied.

There are two initial steps that may be taken along this path. First, as a general framework, it may be useful to tie a narrative-based mediating representation to a rationalistic model like the argumentation approach presented here. One of the problems with the argumentation approach in isolation was that it was based around the idea that people would decompose a problem into bits of evidence, and evaluate these pieces of evidence independently. This turned out not to be an entirely accurate assumption for my domain, as the *connections* between pieces of evidence were at least as important as the bits themselves. Unfortunately, the model provided no ready means for explicitly representing those connections.

Had the interface been able to represent inferences and story elements that were actually important to decision making, the rationalistic / evaluative tools could perhaps have been better used to help evaluate components of the story, and determine which story among possible story proposals was the best fit for the data.

Support may be found for such a model in several places. Pennington & Hastie's work suggests a similar decomposition between story construction and the subsequent evaluation of the constructed stories against verdict criteria. More importantly, there were instances in the collected data from my experiment where the pieces of a correct story were either forgotten or ignored in favor of stories that simply had more information. Such mistakes, or breakdowns in communication, are similar to the breakdowns described by Alterman et al. (2001) that might be examined en route to developing coordinating representations. This is important, as any developed structure should not merely be a good fit for the decision process, but should also help users distribute some of the information they have a hard time managing into the system itself.

In addition to offering some sort of causal representation in the interface, there are modifications to the existing argument representation which may be of some use. Many of these were discussed in Chapter 6. However, the results regarding areas of focused agreement in the non-adaptive groups suggests that some sort of tool to help users to focus their collaborative attention upon important pieces of the story might be of use. Anecdotally, many users complained that it was difficult to know when new information was being posted in the network. Highlighting, or color coding the network to indicate regions of increased activity might help users “come together” around important ideas.

Within the proposed framework, it will be important to keep in mind the preceding observations regarding different types of decision problems, and to choose a decision problem which has similar properties to the murder mystery studied here. One aspect of the work on decision-making presented in this dissertation that has given rise to many of the more interesting findings was the choice of a decision problem that combined different aspects of decision making. This was luck; the problem was chosen because it seemed more compelling, and a better fit for the argumentation formalism that had been implemented. Had the information contained within the problem not exhibited interesting connectedness, this dissertation may have been a lot shorter, but many opportunities would have been left undiscovered.

3 Conclusion

This work began as work to support the tool builder – specifically, the tool builder interested in building adaptive groupware. To this end, I have offered an approach that might be used in some cases to solve a knowledge acquisition problem in groupware. I have also presented a novel target for adaptive functionality, which is the collaborative interaction itself, and have shown how an idealized model of interaction might be leveraged to adapt the collaborative process.

Along the way, there were some unexpected findings about group decision making, which will serve as rich fodder for future investigation. The richness of these results, and the design ideas they support, lead me to offer one final observation about methodology in the design of adaptive groupware.

Part of my effort in the second case study described in the dissertation was to illustrate a technique that worked. My approach, in the abstract, was to define a criteria for success that I had a good chance of attaining. This “good chance” rested upon my confidence that the common knowledge problem was a repeatable problem, and that the adaptive mediation I had developed would enforce the desired outcome, given some reasonably safe bets about people's ability to evaluate information accurately and use the provided formalism properly.

An alternate approach might have been to study group decision making without any mediating structure, adaptive or otherwise, analyze use, develop a mediating structure on top of that, and finally, implement some adaptive functionality. It is unclear that this approach would have led to the same quality of insight afforded by my current analysis. It certainly would have taken longer.

It is generally taken for granted in software design that the first design probably won't be the right one, and that some iteration will necessarily occur before a usable piece of software may be generated. The path I've taken does not eliminate the need for iteration in design. However, it may allow for fewer cycles of iteration. In hindsight, I consider two aspects to my development process to be especially important in this regard.

First, none of my observations would have been possible if not for the fact that it was possible to log incredible amounts of data. Furthermore, the very simple structure that was added to the interface – multiple threads in conversation, and argument tagging – drastically simplified my analysis. The identification of shared consensual focus was a direct result of this interface structure. From this perspective, interface structure is not only a resource for the user, but also for the designer.

The second aspect was that I chose, and enforced, an idealized model of decision making as a the basis for adaptive mediation. This model was well supported in the literature, and represents one scientific community's opinion about what collaborative decision making should look like. Comparing groups with adaptive mediation to those without has not only highlighted differences between the idealized version of what should go on in decision making and what actually *does* go on, but has led to a series of insights about how aspects of either might be combined in the next iteration of design.

There are other idealized models of group processes that deserve similar investigation. I hope that the work presented in this dissertation can serve as a template for others, and can be employed to help further advance the science of design for collaboration.

Appendix A

The following sections contain the murder mystery and all instructional materials used in the second case study. The materials were adapted from the experiment described in Stasser and Stewart (1992). The original materials were authored by Garold Stasser, and revised by Gwen Wittenbaum and Dennis Stewart. They have been revised again for the case study described in this dissertation.

The information about the murder mystery was made available as a single web page with no hyperlinks. The following section presents the entire set of information distributed to participants. Unshared pieces of information are colored according to which participant received that information. Red clues were provided to EE (the Eddie expert), and each them implicate Eddie. Blue clues were provided to BE (the Billy expert), and each of them exonerate Billy. Finally, green clues were provided to ME (the Mickey expert), and each of them exonerate Mickey. Descriptions of the shared clues and differences between the materials used here and the original materials are provided following the mystery

1 The Murder Mystery

MAJOR CHARACTERS

Robert Gill:	The victim
Mary Gill:	The victim's wife
Lt. Mark Moody:	Detective in charge of the investigation
Sgt. Cassini:	Police officer assisting in the investigation
****Eddie Sullivan**:	Handyman who worked for the Gills
****Billy Prentice**:	Yardman who worked for the Gills
****Mickey Malone**:	Owner of MM Auto Parts; business associate of the victim
Sam Nietzel:	Parts manager for Gill Lincoln/Mercury
Dave Daniels:	Owner of Dave's Quick Stop in the Eastwood Shopping Center

** The **ONLY SUSPECTS** under consideration are:

Mickey Malone
Billy Prentice
Eddie Sullivan

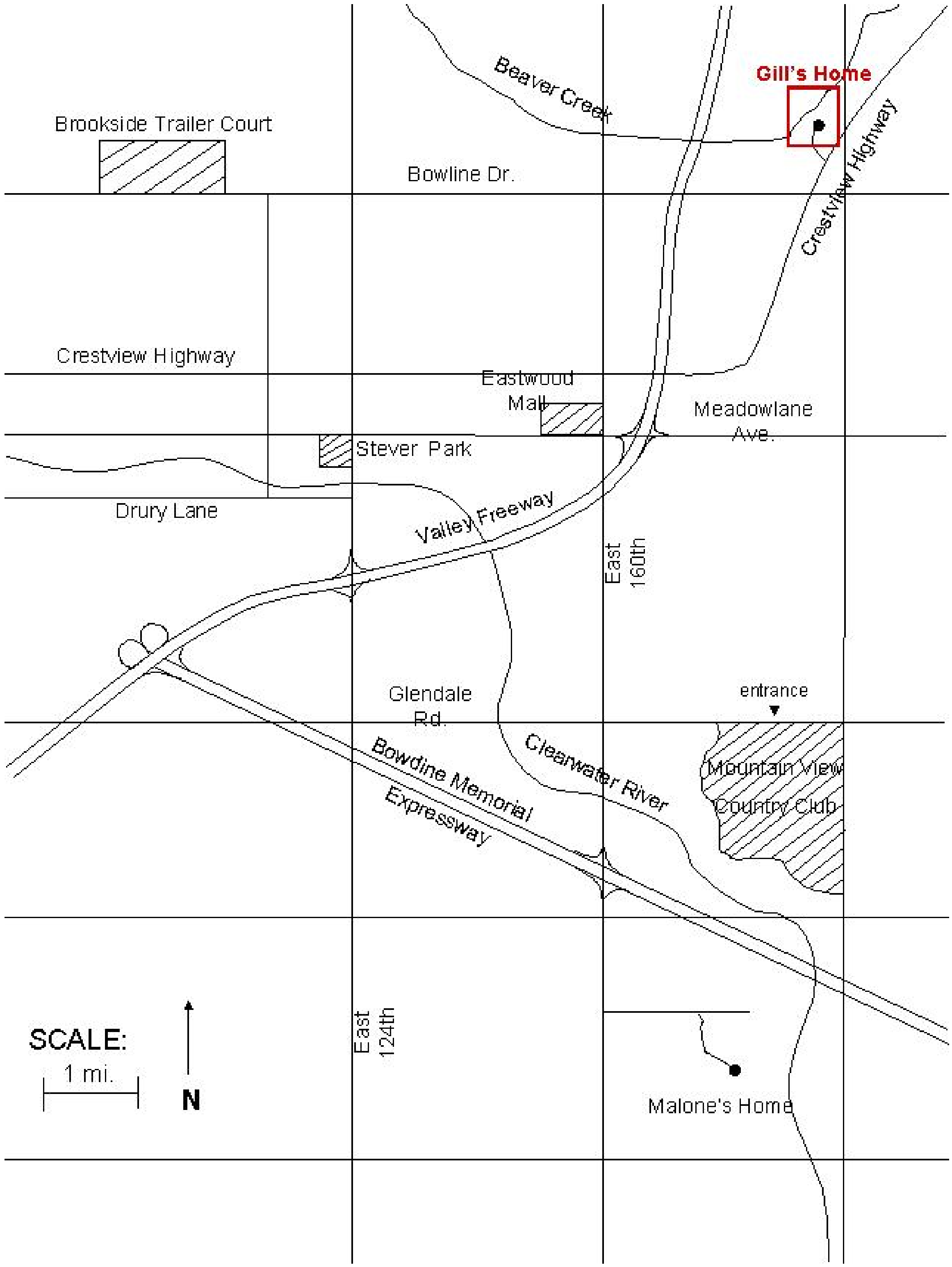


Figure 47: Map of the area

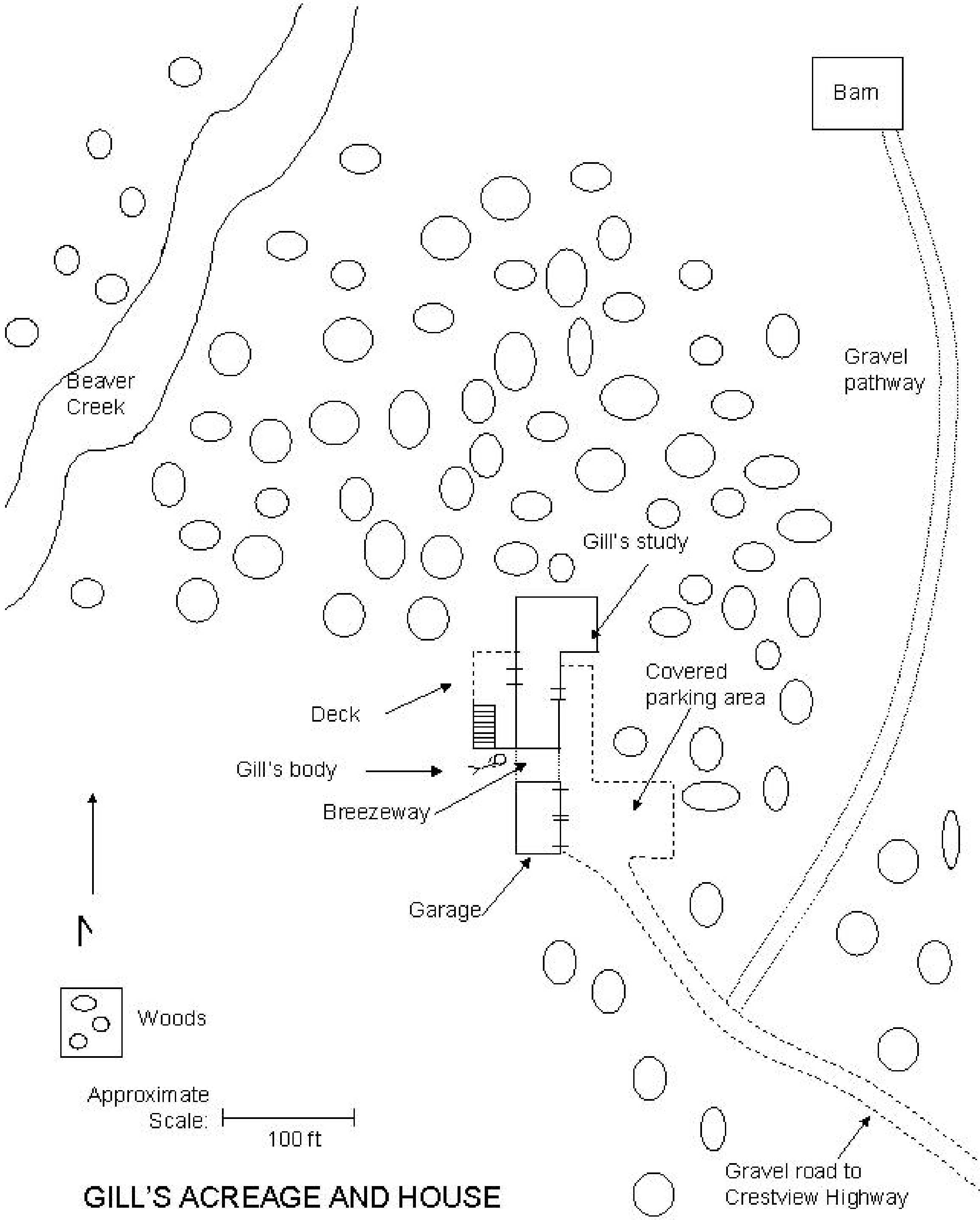


Figure 48: Map of the Gill residence

Excerpts from From

an article in the weekend edition of the local paper, the Valley Sentinel.

Robert Gill, a prominent local businessman was found dead behind his Crestview home this morning. Detective Lt. Mark Moody of the Hilltown precinct reported that Mr. Gill had apparently been assaulted when leaving his home to play golf early this morning. He was struck on the head over the left eye and fell down a flight of stairs leading from a second story deck at the rear of the house.

The preliminary coroner's report concluded that death was caused by injuries sustained from the fall and not from the blow to the head. The report estimated that Mr. Gill's death occurred between 6:30 and 7:00 AM. Lt. Moody would neither confirm nor deny rumors that Mr. Gill had been robbed. "We're following all leads. That's all I have to say for now," said Lt. Moody.

Excerpts from

Lt. Moody's (Lt.M) interview with

Mary Gill (Ms. G), wife of Robert Gill

Lt. M: Mrs. Gill, I know this isn't going to be easy, but I need you to answer some questions for me.

Ms. G: It's okay. It has to be done. And please call me Mary.

Lt. M: Okay,… Mary,… tell me what you remember about Saturday morning.

Ms. G: Well, I always sleep in on Saturdays. I got up around 9 or 9:15 and did aerobics from 9:30 to 10. After that I showered and was drying my hair when I heard a knock at the patio door. It was Eddie Sullivan, our handyman.

Lt. M: That was about 10:30?

Ms. G: Yes, I think so… I'm not absolutely sure.

Lt. M: You are sure about the times that you got up and did aerobics?

Ms. G: Yes, I'm fairly sure of those times. You see, I watch an aerobics program on TV; it's on every morning from 9:30 to 10.

Lt. M: So, when Mr. Sullivan told you that your husband was hurt, what did you do?

Ms. G: He did not actually tell me that it was Bob… not at first. He just said that there had been an accident and that I should call an ambulance… I remember feeling scared, but it didn't occur to me that it might be Bob.

Lt. M: This was the first time that you knew that something had happened?

Ms. G: Yes, Bob always plays golf on Saturday morning; he always leaves early and doesn't return until 11 or so. I thought he was at the country club.

Lt. M: You said that you were scared. Did you suspect the Mr. Sullivan wasn't telling you everything?

Ms. G: Not really… I suppose that I was just reacting to the urgency in Eddie's voice.

Lt. M: Do you know if Mr. Gill ever left the house during the morning?

Ms. G: No, I'm not sure. All I remember is that he was talking on the phone in the study… it's across the hall from the bedroom. I remember it was light outside... must have been around 6. Next thing I knew, I heard voices, or a voice… shouting… I'm not sure. I was still half-asleep. It sounded like it was coming from outside.

Lt. M: Where is the bedroom located?

Ms. G: At the back of the house, on the northwest corner. Anyway, I thought it was Bob. I thought maybe he was yelling at the cat. She sometimes runs out the patio door when he's leaving and it infuriates him. But then I heard what sounded like a groan… and something fall. This woke me up completely. I went to the window and looked out but didn't see anything.

Lt. M: Can you see the deck from your window?

Ms. G: No, not very well. I remember looking at the clock. It was about 6:40. I thought, "Bob is usually gone by now." Then I heard a car on the gravel driveway. I went to the study window at the front of the house but didn't see anyone.

Lt. M: Did you think it was a car leaving?

Ms. G: Yes, I thought so. I saw Bob's pickup in the carport and I assumed that he took the Mercury out of the garage. Sometimes he takes the pickup. I remember thinking that the noise I heard must have been the garage door closing. It always comes down with such a bang.

Lt. M: Can you see under the carport from the study window?

Ms. G: Oh… yes, I can see under it completely.

Lt. M: So you suspected nothing until Mr. Sullivan came to the patio door?

Ms. G: That's right. I thought it was unusual that he was at the back door. He usually comes to the front door. And he looked upset. He opened the door partway when he saw me and shouted, "Call an ambulance. There's been an accident." Or something like that. He made it sound very urgent.

Lt. M: So what did you do?

Ms. G: I called for an ambulance like he said… Then I went out… on the deck (bursts into tears).

Lt. M: I know that this is hard Mary… but we're just about through.

Ms. G: I can't go on.

Lt. M: I know you are upset, but please try to continue. It's very important… now when did you realize that it was your husband?

Ms. G: When I got out on the deck… I looked down over the railing… I was stunned… Eddie looked up at me and shook his head. Then I could tell that it was Bob, and somehow I knew that he was… gone (sobbing). I fell apart. I couldn't stand looking at him. I went back inside…I stayed there until the ambulance came. Eddie came up, and asked if there was anything he could do. I think I asked him to call my sister. Anyway she got here just before the ambulance.

Lt. M: Mary, there was no wallet or identification on your husband. Did he ordinarily carry a wallet?

Ms. G: Yes, he always does.

Lt. M: Did he carry a lot of money?

Ms. G: Not a large amount. Usually no more than $50.

Lt. M: Do you mind if we see if he left his wallet somewhere in the house on Saturday?

Ms. G: No, go ahead.

Lt. M: Thank you for your help Mary. Take care.

[Lt. Moody and Mary Gill searched the house, but did not find the wallet.]

Excerpts from
Sgt. Cassini's (Sgt. C) Interview with
Eddie Sullivan (Ed. S), The Handyman

Sgt. C: Mr. Sullivan, you said that you arrived at Mr. Gill's about 6 Sat. morning. You were tearing down a barn for him, I believe.

Ed. S: Yeah.. about 6… the sun was just coming up. I like to get my work done early before it gets real hot.

Sgt. C: Did you notice anything unusual when you arrived?

Ed. S: No… The light was on in Mr. Gill's study, but that wasn't unusual. He is always up when I get there in the morning. He was a hard worker. He earned his money; it wasn't given to him.

Sgt. C: How did you happen to notice Mr. Gill's body?

Ed. S: I went back to my truck to get my crowbar. I left it laying next to the truck. When I got there, the crowbar was gone. I looked around… that's when I saw Mr. Gill laying in the grass through the breezeway. At first, I thought it was Billy..

you know Billy… ah … Prentice, he cuts the grass on Saturdays. He's always there bright and early and I though maybe he had hurt himself. Anyway, I ran back there. I was shocked to see Mr. Gill. I didn't think he was even there 'cause he plays golf on Saturday morning. He leaves at 6:30, regular as clockwork, and is never back till about noon.

Sgt. C: Okay, so you ran over to Mr. Gill…

Ed. S: Yeah, like I said I was shocked. He looked real bad… blood on his head and laying there real awkward. I ran up the stairs and pounded on the patio door. I started to open it and then I saw Mrs. Gill coming in from the living room. I thought I shouldn't alarm her too much so I just said, "Call an ambulance. There's been an accident." She started to run past me like she knew it was bad but I stopped her and said, "It's alright, just call the ambulance." I never told her it was Mr. Gill. I didn't know he was dead till I got back down the stairs.

Sgt. C: Did you ever find the crowbar?

Ed. S: What?… Oh… no. I never did. I never looked again. I was real upset. I didn't even go back to the barn. I just left after the ambulance came. By that time, Mrs. Gill's sister and her husband were there. I didn't figure that I could do anything.

Sgt. C: You said at first you thought it was Billy Prentice lying there in the grass instead of Mr. Gill. Was Billy there Saturday morning?

Ed. S: You know I don't know… come to think of it his car wasn't there and none of the yard tools, or the lawn mower, was out. But I thought I heard his station wagon earlier.

Sgt. C: When was that?

Ed. S: I can't say for sure. I just remember hearing a car with a loud muffler and thinking, "That's Billy." None of Gill's cars would ever sound like that. I'd guess around 7.

Sgt. C: Did you hear anything else? Did you hear anything like a fight or, perhaps, Mr. Gill falling?

Ed. S: No, can't say I did. You know the barn is quite a ways from the house… probably 200 or 300 yards. And there's woods between there too.

Sgt. C: You said you went back to pick up your crowbar by your truck. Where was your truck?

Ed. S: It was in the carport beside Gill's pickup.

Sgt. C: Why didn't you drive it down to the barn where you were working?

Ed. S: Well… it had rained the night before, and I didn't want to get it stuck down there. There's a gravel path but it's not wide enough. Besides Mr. Gill didn't want me making ruts in the grass.

Sgt. C: Eddie, did you and Mr. Gill get along?

Ed. S: Yeah… I always liked him… He was real fair when it came to business… paid well… easy to work for.

Sgt. C: Your daughter worked at Gill's car dealership, didn't she? How did they get along?

Ed. S: Yeah… She was his bookkeeper for several years. All of a sudden she quit. I didn't ask her about it. She seemed upset, but I figured that that was their business. You know what I mean?

Sgt. C: Sure, if you think of anything else that I should know, give me a call. I'll be in touch.

Excerpts from
Lt. Moody's (Lt. M) Interview with
Mickey Malone (M.M.), owner of MM Auto Parts

Lt. M: Mr. Malone, I have to ask you some hard questions. It's well known that you and Mr. Gill go back a long way but things were kind of rough between the two of you lately.

M. M.: We had some differences.

Lt. M: Did you call Mr. Gill Saturday morning?

M. M.: Yes

Lt. M: Why?

M. M.: Well… we always play golf with two other fellows on Saturday mornings… a foursome, you know. Well… the last 2 weeks things had been awkward… downright nasty at times. I told him we either put this thing behind us or else… or else either he or I should drop out of the foursome. It just wasn't fair to the others… to ruin their golf.

Lt. M: The other 2… Rick Rooney and Jim Townsend, I believe.

M. M.: Yeah. Anyways, I wanted to clear things up before we got to the country club.

Lt. M: You play at Mountain View?

M. M.: Yeah.

Lt. M: What did Mr. Gill say when you called him?

M. M.: Bob told me to stuff it. I told him, "If you're playing golf, I'm not!" He said, "Fine, do what you want."

Lt. M: What did you do?

M. M.: My first impulse was to drive over to his place and work this out face-to-face. I go to the Crestview turnoff and thought to myself, "This is silly. We'll just end up fighting," so I turned around and headed straight to the golf course. Just 'cause Bob wanted to be a horse's rear didn't mean I had to ruin my day.

Lt. M: What time did you leave home?

Lt. M: What time did you leave home?

M. M.: 6:30 ... 6:40... I don’t know… somewhere around then.

Lt. M: How long does it take to get to Crestview from your house?

M. M.: I don’t know. Maybe… it’s about a mile north of Meadowlane… that’s about 9 miles then… probably 15 minutes

Lt. M: And then you turned around and went to the golf course?

M. M.: Right.

Lt. M: It’s about 6 miles from Crestview to Mountain View golf course?

M. M.: Yeah… about.

Lt. M: So you left home somewhere around 6:30 or 6:40. Fifteen minutes to Crestview and perhaps another 10 minutes back to Mountain view… Let’s see… That should have put you at the golf course around 7, give or take 5 minutes. Is that about right?

M. M.: Sounds right.. yeah, I got there a bit before 7, actually, when we usually meet.

Lt. M: You did not go to Gill’s place on Saturday morning?

M. M.: No, I didn’t.

Excerpts from
Lt. Moody’s (Lt.M) interview with
Billy Prentice (B.P.), Gill’s Yardman

Lt. M: Billy, I need to talk to you about Mr. Gill’s death. You did hear about it didn’t you?

B. P.: Yes, sir. It was awful, wasn’t it?

Lt. M: Yeah, too bad. Were you at Mr. Gill's place on Saturday morning?

B. P.: No, sir.

Lt. M: Don't you usually cut the grass on Saturday?

B. P.: Yes, sir… usually…, but not last Saturday.

Lt. M: Why not?

B. P.: Ah.. I just cut it the week before.

Lt. M: But this time of year.. don't you usually cut it every week?

B. P.: Yeah, but … I wasn't feeling good last Saturday morning.. Besides it rained Friday night and the grass was probably wet.

Lt. M: But I mowed my grass last Saturday morning. By 9:30, the sun had dried the grass out. Remember, it was clear and hot. Didn't it occur to you that the grass would be dry later in the morning?

B. P.: I guess so… but by then I figured I wouldn't have time to get it done before my ball game.

Lt. M: Billy, what time was your ball game?

B. P.: Noon.

Lt. M: How long does it usually take you to cut the grass?

B. P.: A couple of hours, but I had other things I needed to do out there.

Lt. M: Couldn't you have done those other things while the grass was drying and still been able to make it to your ball game?

B. P.: I suppose so… l don't know… I like to get to the game early… Besides I said I wasn't feeling so good in the morning.

Lt. M: Billy, I should tell you right out… Mr. Sullivan… You know Mr. Sullivan don't you?

B. P.: You mean Eddie, the carpenter? Yes, sir, I know him.

Lt. M: Well, Mr. Sullivan heard your car at Gill's on Saturday morning. How do you explain that?

B. P.: How'd he know it was my car? When?

Lt. M: He just said that he heard your car about 7 Saturday morning. He said that he recognized the loud muffler.

B. P.: No, he couldn't have. I wasn't there at 7 on Saturday.

Lt. M: Billy, come one. We know that your car was at Gill's place. We picked up fresh tire tracks along the edge of the gravel near the carport. They match your tires, Billy, and we know they weren't a week old.

B. P.: Okay… Okay… I was there Friday to ask Mr. Gill for an advance. I was a little short on money. He gave it to me.

Lt. M: What time on Friday?

B. P.: Around 4:00, just before ball practice. I was broke and he always helps me out.

Lt. M: So you borrow money a lot? What do you need the money for?

B. P.: Ah… yeah, I suppose so… for my car. I work on my car a lot, fixing it up, keeping it in good shape.

Lt. M: Okay, Billy, that's all for now. We'll talk later.

B. P.: Sir,… you know I didn't hit Mr. Gill… You know I wouldn't hurt him… He was always good to me.

Lt. M: Sure, Billy, I know… See you around.

Excerpts from
Lt. Moody's (Lt. M) Interview with
Rick Rooney (R.R.), Gill's and Malone's Golf Partner

Lt. M.: I'd like to ask some questions about Mickey Malone.

R. R.: I'll be glad to help if I can.

Lt. M.: You play golf with Mr. Malone on Saturday morning. Right?

R. R.: Yes, I do. We have a regular foursome.

Lt. M.: Can you tell me anything about his relationship with Mr. Gill?

R. R.: They were always good friends… until these last few weeks. They had some sort of business disagreement. Mickey wouldn't say a whole lot about it, though. They've had problems in the past, but it's never been this bad.

ITEM 61: Lt. M.: What time did Mr. Malone arrive at the golf course last Saturday?

R. R.: Well, I got there around 7, but he was already there and had picked up our golf cart. So I'm not exactly sure, but it must've been at least 10 minutes before 7. There's always a line for carts on Saturday morning.

Lt. M.: Okay, I appreciate your help.

Excerpts from
Sgt. Cassini's (Sgt. C) Interview with
Dave Daniels (D. D.), Owner of Dave's Quick Stop
In the Eastwood Shopping Center

Sgt. C: Dave, when you called Saturday morning, you said that you found a wallet behind your store. Where did you find it?

D. D.: It was laying beside the dumpster in the back… next to some boxes that I had stacked out there.

Sgt. C: What did the wallet look like?

D. D.: It was a nice one. It looked new… and expensive… so I thought it was strange that someone would throw it away.

Sgt. C: Did it have any money in it?

D. D.: No, in fact it was empty. All that I found were Mr. Gill's credit cards inside the dumpster.

Sgt. C: You never found any money or a driver's license?

D. D.: No, just 3 credit cards.

ITEM 62: Sgt. C: What time did you find the wallet?

D. D.: Probably about 7 AM. Yeah, I remember because I got to the store just before 7 and was checking some stock in the back room right before I found the wallet.

Sgt. C: What made you go outside?

D. D.: I heard a car pull up in back and then speed away. I went out to see what was going on but the car was gone by the time I got out there. That's when I saw the wallet.

ITEM 43: Sgt. C: So you heard a car right before you went out and saw the wallet. Are you sure it was a car?

D. D.: No… not really. I assumed it was a car because it ran real quiet. I probably wouldn't have heard it but the tires squealed when it left. Like I said, I didn't see it.

Sgt. C: Could it have been a pickup?

D. D.: I suppose so.

Sgt. C: Are you sure the wallet wasn't there earlier?

D. D.: Pretty sure, I walked right past there when I came in just a little earlier and I don't know how I would have missed it if it was there.

Sgt. C: Thanks, Dave. If you think of anything else, call me.

Excerpts from
Lt. Moody's (Lt. M) Interview with
Sam Nietzel (S. N.), Parts Manager for Gill Lincoln/Mercury

Lt. M: Mr. Nietzel, I need to ask you some question in regards to Gill's connection with MM Auto parts. Were Gill and Malone having difficulties?

S. N.: Yes, I suppose so. We've done business with Malone for years. In fact, he started supplying parts for us when he was still operating out of the barn on the old Malone place.

Lt. M: I've heard that Malone got his real start by being a supplier for Gill and they were friends for years.

S. N.: Yes, that's right… They've been friends way back… but they had their ups and downs… They always worked things out before… until this last thing. It seems that Malone started giving Gill substandard parts, which really steamed Bob because he is very concerned about giving his customers quality service. He even told me to stop ordering from Malone.

Lt. M: What was wrong with the parts?

S. N.: Well, some of them didn't fit; some seemed to wear out and break easily. My guess is that they were either rebuilt or after market parts.

Lt. M: When did you realize this was going on?

S. N.: About 2 months ago… It's been a mess around here since.

Lt. M: So would you like to go back to MM Parts?

S. N.: No, and I especially wouldn't do it without Mr. Gill's 'Okay'. And he was dead-set against it... Wouldn't even talk about it! Mr. Gill was a proud and stubborn man.

Lt. M: Another matter… Do you know anything about Ms. Sullivan's leaving the firm?

S. N.: That would be Sue Sullivan, the bookkeeper?

Lt. M: Yes, Sue Sullivan.

S. N.: No, I don't really know anything in particular.

ITEM 11: Lt. M: Nothing out of the ordinary happened before she left?

S. N.: Well, maybe… I didn't know there was a problem until I overheard them arguing in his office. I didn't mean to hear, but I couldn't help it. I was going in to talk to Mr. Gill about something. Next thing I knew she was leaving… I mean leaving for good, packing up her things.

Lt. M: Did you hear what they were arguing about?

S. N.: No, they sounded real mad, but I couldn't make out what they were saying… They stopped when they saw me coming.

Lt. M: Had they argued like that before?

S. N.: No, not that I know of. They always seemed to get along real well. Maybe … well, I don't know. Mr. Gill was a stand-up guy.

Lt. M: Do you know if Mr. Gill had any enemies or dissatisfied customers?

S. N.: Not really. Mr. Gill treated his customers like royalty. He always said, "The customer is always right… always!" He not only said it, he lived by it.

Lt. M: Thanks, Sam, for your time. If you think of anything else, give me a call. You've been a great help.

S. N.: Glad to be… Want to get this mess sorted out.

Excerpts from
Lt. Moody's (Lt. M) Follow-up Interview with
Mary Gill (Ms. G), wife of Robert Gill

Lt. M: Mary, I need your help to clear up a couple of matters if you don't mind.

Ms. G: Sure.

ITEM 41: Lt. M: Billy Prentice claims that he came by on Friday to ask your husband for an advance. Do you know anything about this?

Ms. G: Why… yes, he did come by… in the afternoon, I believe.

Lt. M: Did your husband give him any money?

Ms. G: Yes, he did. I'm not sure how much, but I remember he said, kind of jokingly, "I wonder if I'll ever get to pay Billy AFTER he does the work."

Lt. M: Do you know exactly what time it was on Friday when he came by?

Ms. G: No, I'd only be guessing… late afternoon, I'd say.

Lt. M: Billy Prentice seems to have problems handling his money. Does he borrow… or ask for advances on his wages… often?

Ms. G: Yes… quite often.

Lt. M: Do you have any idea what he uses the money for?

Ms. G: Well… I'm not sure, but I think he has been involved with gambling.

Lt. M: What makes you say that?

Ms. G: Well, I know he plays poker with some friends of his, and Bob and I ran into him once at the racetrack. We've only been there a few times, but I always like to go just to watch the horses. I think they're beautiful. Anyway, Bob and I never bet more than a few dollars. But when we saw Billy there, he had quite a stack of betting slips in his hand. He noticed us just then and seemed really nervous and quickly walked away from the betting window. After this incident, Bob said he would keep an eye on Billy.

Lt. M: How long ago did this happen?

Ms. G: Hmm… about… it was soon after he started working for us. Probably 2 years ago.

Lt. M: One other thing, was Billy here anytime on Saturday morning?

Ms. G: No, I can't say that he was, come to think of it. I guess with everything else I never gave it a thought, but he didn't show up… At least, he never mowed the lawn.

ITEM 13: Lt. M: As I recall, you heard a car on the gravel out front about 6:40. You thought at the time that it was your husband driving away. Could it have been Billy or someone else driving up the drive?

Ms. G: Maybe… but, no. It couldn't have been anyone driving up… If it had been it seems that I would have seen them. The only thing that I saw was Bob's pickup in the carport… nothing else.

Lt. M: Did Mickey Malone come by anytime Saturday?

Ms. G: No… I don't think so… Melissa, Mickey's wife, called early Saturday afternoon. She said that they had just heard on the radio and wanted to know if there was anything that they could do.

Lt. M: They didn't come over at any time?

Ms. G: Not on Saturday. They stopped by briefly on Sunday to offer their condolences.

ITEM 12: Lt. M: One other matter… Is it true that Eddie Sullivan has a hearing problem?

Ms. G: Yes, he is very hard of hearing. Sometimes when he gets a phone call, I have to call him. I've tried calling to him from the deck, but he never hears me. I have to walk right up to him before I can get his attention.

Lt. M: Doesn't he have a hearing aid?

Ms. G: He has one, but he doesn't wear it while he is working. He says that it doesn't fit well. It's one of those tiny ones and he's afraid he will lose it.

Lt. M: I think that's all, Mary. Thanks for your patience. I hope I don't have to bother you again with these details.

Excerpts from
Sgt. Cassini's (Sgt. C) Follow-up Interview with
Billy Prentice (B. P.), Gill's Yardman

Sgt. C: Billy, since you talked with Lt. Moody, some new things have come up. I remind you, Billy, that you don't have to answer my questions if you don't want to.

B. P.: Sir, I don't mind. I have nothing to hide.

Sgt. C: Very well. You said that you went to Gill's on Friday night, not Saturday morning… Right?

B. P.: Yes, sir… Well, actually I went there Friday afternoon, not Friday night.

Sgt. C: To borrow some money, I think you said.

B. P.: Yes, that's right.

Sgt. C: Was this money to pay off gambling debts, Billy?

B. P.: No. No, sir.

Sgt. C: Is it true that you are an excessive gambler?

B. P.: No! I mean.. well, I gamble as much as the next guy… you know, poker with the boys… racetrack every now and then. I used to do it a lot more a couple years back, but I've really cut down. I don't have a problem with it sir… really!

Sgt. C: Okay, so you were there on Friday and not on Saturday?

B. P.: Yes, sir, that 's what I said!

Sgt. C: Those tire tracks that Lt. Moody told you about… Billy, those tracks were almost certainly made after Friday night's rain. And, as you know, it rained from about 10 to midnight.

B. P.: But… [long pause]… I…

Sgt. C: Billy are you sure that there is nothing that you want to tell me?

B. P.: Alright, sir… I was there… I went to do some work. I saw Mr. Gill just laying there. I went over to him. It was awful.

Sgt. C: Billy, why didn't you say something before?

B. P.: Nobody's going to believe me. I thought I'd just better get out of there and act like I didn't know nothing.

Sgt. C: So you ran.

B. P.: I sure did. I almost hit Mr. Sullivan's truck when I was pulling out of the carport. I couldn't get out of there fast enough. I swerved… that's probably when I went off the road.

Sgt. C: While you were at Gill's did you see a crowbar?

B. P.: What? A crowbar? .. yeah, I did now that you mention it.

ITEM 42: Sgt. C: Where did you see it?

B. P.: It was laying in front of the garage door, the side door where I get the mower out. I remember moving it to the side so I could get the mower out.

Sgt. C: That's all? You just moved it over to the side?

B. P.: Yes, sir.

Sgt. C: Anything else you remember about that?

B. P.: Well, I remember thinking that it must be Mr. Sullivan's 'cause Mr. Gill didn't have any tools to speak of around the place… excepting some garden tools. But then I thought that that was odd 'cause Mr. Sullivan always makes this big thing about keeping his tools locked up when he's not around. And I didn't see him anywhere… just his truck.

Sgt. C: Billy, the crowbar was found in the bushes south of the garage… with your fingerprints on it. Can you explain that?

B. P.: No, Sarg, I swear… If I did pick it up, why would I throw it in the bushes?

Sgt. C: That's what I'd like to know… Okay, let's go on. What time would you say that you were at Gill's on Saturday?

B. P.: I don't rightly recall. It was late. Maybe 8, I'd guess. Like I said, I wasn't feeling so good.

Sgt. C: Did you take Mr. Gill's billfold?

B. P.: No, sir. You gotta believe me. When I saw he was dead, I just got out of there!

Sgt. C: How did you know that he was dead?

B. P.: I don't know… He looked dead… He didn't move when I yelled… he wasn't laying there natural like.

Sgt. C: Did you go over to him? Did you check his pulse? Didn't you even try to get help? Maybe call Mrs. Gill or something?

B. P.: No, I just got out of there. I didn't think there was anything I could do.

Sgt. C: Okay, thanks Billy. That'll be all for now.

Excerpts from
Sgt. Cassini's (Sgt. C) Follow-up Interview with
Eddie Sullivan (Ed. S), the Handyman

Sgt. C.: Eddie, since we last talked we found your crowbar in the bushes south of Gill's garage. At least, we think it's yours: it has "ES" stamped on it.

Ed. S.: Yeah, all of my tools are stamped. You can't be too careful. People borrow them and forget they're yours. You know what I mean?

Sgt. C.: Do you have any idea how it got in the bushes?

Ed. S.: No… Can't say I do.

Sgt. C.: We're trying to get some things about last Saturday sorted out. You said you got to the Gill's about 6 in the morning and went straight to the barn. Then about 7 you heard a car… with a loud muffler. Mrs. Gill thought you came to the patio door around 10:30. Is that about the time that you discovered Mr. Gill's body?

Ed. S.: I'm not sure about that. It could have been around then. I really don't remember.

Sgt. C.: Okay, Eddie, if you think of anything else, give me a call.

Excerpts from

Lt. Mood's (Lt. M) Follow-up Interview with
Mickey Malone (M. M.), owner of MM Auto Parts

Lt. M.: Mr. Malone, I need to double-check some things that you told me the other day. Did Mr. Gill write this note to you?

[Copy of note on following page.]

M. M.: Ah, yeah, he did.

Lt. M.: When did you receive the note?

M. M.: I think it was last Thursday. When I read that he was going to ruin my business with other customers, I offered him the best terms I could. I was even willing to sell him parts at cost to try and patch things up. I mean… if I lost the business, I don't know what I would do. But he can be so stubborn. That's why I called him Saturday morning. I just thought it was time to work this out, one on one.

Lt. M.: Well, is it true that you gave defective parts to Mr. Gill?

M. M.: I'd never knowingly give Bob bad parts. But I have recently hired a part-time sales guy – Louie Brown - to help out. He has also been doing some of the purchasing for me.

Lt. M.: Is it possible Mr. Brown could be using a different supplier without your knowledge?

M. M.: That actually occurred to me after I cooled down a little on Saturday. I placed a call to him after the golf game to try to figure it out, but I got his machine. Still haven't heard from him.

Lt. M.: What time did you call him?

M. M.: Right after our golf game – around noon.

Lt. M.: I see. Okay, Mr. Malone, that's all for now.

GILL LINCOLN / MERCURY

From the desk of

ROBERT GILL

President

Mickey,

I am very upset about the substandard parts I have been receiving from you. I know we've had our problems in the past, but I never thought you would go this far. I am a man of integrity and will not tolerate such maneuverings from business colleagues. Needless to say, I will have to notify my customers and other dealers about the quality of MM auto parts.

Bob

ITEM 63: Louie Brown

Excerpts from
Lt. Moody's (Lt.M) Interview with
Louie Brown (L.B.), part-time accountant for MM Auto Parts

Lt. M: I'd like to ask you a few questions if you don't mind.

L. B.: Sure. What can I do for you?

Lt. M: I understand that you do some work for Mickey Malone's auto-parts supply?

M. S.: That's right. I've been doing a bit of this and that for MM Auto Parts for several weeks now.

Lt. M: And have you been able to help the business out?

L. B.: Well, I'm happy to say I think I've been able to improve his profit margins.

M. S.: Where were you last Saturday?

L. B.: The wife and I took the kids to the amusement park.

Lt. M: Did you receive any messages while you were gone?

L. B.: As a matter of fact, I did. From Mr. Malone. He wanted to know if we had recently changed any of our suppliers. Mentioned something about one of his larger clients complaining about sub-standard parts, and he wanted to get it cleared up.

Lt. M: And... have you changed suppliers?

L. B.: Look, what's this all about, now? Am I under arrest or something?

Lt. M: Please relax Mr. Brown, I'm merely trying to corroborate some information. Do you know what time Mr. Malone called?

L. B.: Okay, well, it was around noon – the message on the answering machine has a timestamp. It's still there if you want to hear it.

Lt. M: That won't be necessary. Thanks for your time.

1.1 Evidence Summary

EDDIE

<u>Implicating</u>:

IE1.11 Sue S's argument with Mr. Gill and possible affair (SN)
IE2.12 Hearing problem (MsG2)
IE3.13 Mary did not see Eddie's truck in carport at 6:40 (MsG2)
IE4.14 Eddie habitually locks up tools (BP2)
IE5.15 Eddie knew that Gill always left around 6:30 (ES1)
IE6.16 Left his crowbar out for over several hours (ES1, ES2)

BILLY

<u>Implicating</u>:

IB1.31 Problems with money and gambling (MsG2, BP2)
IB2.32 No wallet on body; wallet was later found without money (MsG1, DD)
IB3.33 Eddie reported hearing Billy's car around 7 (ES1, ES2, BP2)
IB4.34 Fingerprints on crowbar (BP2)
IB5.35 Lied about being at Gill's Sat morning (BP1, BP2)
IB6.36 Tire tracks made Sat morning matched Billy's (BP1, BP2)

<u>Exonerating</u>:

EB1.41 Mary confirmed borrowing money (MsG2)
EB2.42 Billy's story about moving crowbar (BP2)
EB3.43 Car that dropped wallet quiet (DD)

MICKEY

<u>Implicating</u>:

IM1.51 Business Feud with Gill (MM1, SN, RR)
IM2.52 Argued with Gill on phone Sat morning (MM1)
IM3.53 Given time left home, would have arrived at Gill's about 6:40 (MM1)
IM4.54 Wallet found near route that Mickey would have taken from Gill's to golf course (MM1, DD)
IM5.55 Note from Gill to Mickey (MM2)
IM6.56 Gill's continued refusal to accept Mickey's offer (MM2, SN)

<u>Exonerating</u>:

EM1.61 Arrived at golf course before 7:00 (RR)
EM2.62 Car dropped wallet at 7:00 (DD)
EM3.63 Still trying to solve disagreement with Gill after Gill's death (LB)

Figure 49: Summary of evidence items

Figure 49 contains a summary of each of the evidence items used in the mystery. As with preceding sections, unshared information items are colored according to the participant that received the clue. The codes in parentheses following each information item indicate where in the mystery the item occurs. For instance "MsG2" indicates Mrs. Gill's second interview.

1.2 Differences from the original

Two modifications were made to the original mystery, based on experiences in the pilot study. First, the unshared clue IE1.11 was reworded slightly to make a slightly stronger implication that perhaps Mr. Gill had done something to Eddie Sullivan's daughter Sue. Only the last statement in the unshared text was changed. The original wording was:

> S. N.: No, not that I know of. They always seemed to get along real well. Maybe it was… well, I don't know.

The modified wording was:

> S. N.: No, not that I know of. They always seemed to get along real well. Maybe … well, I don't know. Mr. Gill was a stand-up guy.

Despite this modification, collaborators in general had a very hard time identifying Eddie's motive.

A second modification was made to reduce the strength of Mickey's alibi. In the original mystery, Mickey stopped on the way to the golf course for a cup of coffee, and this was corroborated by an interview with the waitress at the coffee shop. This interview was unshared information, and was provided instead of the interview with Louie Brown, which was not part of the original text. Several modifications were made to the interviews with Mickey to accommodate these changes, and to the unshared clue

EM1.61.

In the first interview, some changes to the timing of events as described by Mickey were necessary. In the original mystery, the portion of the interview that deals with timing is as follows:

Lt. M: What did you do?

M. M.: My first impulse was to drive over to his place and work this out face-to-face. I go to the Crestview turnoff and thought to myself, "This is silly. We'll just end up fighting." I turned back.. stopped at a coffee ship across from Eastwood… there on 160th. I thought about it some more and decide to go play golf. Just 'cause Bob wanted to be a horse's rear didn't mean I had to ruin my day.

Lt. M: What time did you leave home?

M. M.: 6:20.. 6:30… I don't know… somewhere around then.

Lt. M: How long does it take to get to Crestview from your house?

M. M.: I don't know. Maybe… it's about a mile north of Meadowlane… that's about 9 miles then… probably 15 minutes

Lt. M: How long were you at the café?

M. M.: I don't recall for sure. Why? What does it matter anyway?

Lt. M: Mr. Malone, you and Mr. Gill were not on the best of terms. To be honest, we don't know what happened last Saturday morning but it's clear that there was foul play. We are just following all leads. If you don't want to answer my questions, you don't have to… at least, not right now.

M. M.: I really don't know. As I remember, I drank 2 cups of coffee and then left… maybe 10 minutes.

Lt. M: You went to the golf course straight from the café?

M. M.: Right.

Lt. M: It's about 5 miles from Eastwood to Mountain View golf course?

M. M.: Yeah… about.

Lt. M: So you left home somewhere around 6:20 or 6:30. Fifteen minutes to Crestwiew; a couple of minutes back to the café, let's say; 10 or so minutes of coffee drinking; and say another 8 minutes to Mountain view… Let's see… That should have put you at the golf course around 7, give or take 5 minutes. Is that about right?

M. M.: Sounds right.. yeah, I got there right at 7; that's when we always meet.

Lt. M: You did not go to Gill's place on Saturday morning?

M. M.: No, I didn't.

Lt. M: Thanks, Mr. Malone, for your time.

The unshared information in Rick Rooney's testimony was also changed from the original to corroborate Mickey's statement. The original text was:

Lt. M.: What time did Mr. Malone arrive at the golf course last Saturday?

R. R.: Around 7 as usual.

The waitress' testimony in the original mystery was:

Excerpts from
Lt. Moody's (Lt.M) Interview with
Millie Smith (M.S.), a waitress at Ray's Café

Lt. M: I'd like to ask you a few questions if you don't mind.

M. S.: Sure. What can I do for you?

Lt. M: Were you working here last Saturday?

M. S.: Yes, I was. The morning shift.

Lt. M: So you would have been here at 7 in the morning?

M. S.: Yes, I was covering the counter and the cash register.

Lt. M: Do you remember seeing this man in here last Saturday?

[He shows her a picture of Mickey Malone.]

M. S.: Hmm… yes, as a matter of fact, I do. He came in early; I'd guess around 6:30 or 6:45. Somewhere around then, shortly after I got here. Ordinarily I wouldn't remember him because we get a lot of one-time traffic from the freeway and I know he's not a regular. But I do remember because he just sat there and drank 2 cups of coffee rather quickly and then left all of a sudden - - like he was late for something. He didn't even wait for his bill. He just left 2 dollars on the counter. I remember thinking, "I wish everyone would tip like that for a cup of coffee."

Lt. M: How long was he here?

M. S.: Not very long… maybe 10 minutes, 15 minutes at most.

Lt. M: Can you tell me anything else about him? You said he seemed in a hurry. Did you notice anything else unusual?

M. S.: No, I wasn't paying that much attention to him. We were pretty busy at the time.

Lt. M: Okay, thanks for your time. Here's 2 dollars for the coffee. See Malone isn't the only big tipper. We cops appreciated service with a smile, too.

Finally, the text from Mickey's second interview was somewhat modified from the original in order to develop the story-line around Louie Brown. The original interview is as follows.

Excerpts from
Lt. Moody's (Lt. M) Follow-up Interview with
Mickey Malone (M. M.), owner of MM Auto Parts

Lt. M.: Mr. Malone, I need to double-check some things that you told me the other day. You said you left home about 6:20 or 6:30 with the idea of going to Mr. Gill's house to talk to him.

M. M.: That's right.

Lt. M.: But you never actually went to his place, rather you stopped for coffee.

M. M.: Yes

Lt. M.: So you were at the café about 10 minutes, I believe you said. And then you went to the golf course, arrived there around 7?

M. M.: Yeah.

Lt. M.: You said that you had made overtures to Mr. Gill, wanted to clear things up.

M. M.: That's right, but he's been so hard headed.

Lt. M.: Well, is it true that you gave defective parts to Mr. Gill?

M. M.: My auto parts are of respectable quality as far as I know. If Bob had any problems with the parts I gave him, it wasn't my doing.

Lt. M.: Did he write this note to you?

[Copy of note on following page.]

M. M.: Ah, yeah, he did. When I read that he was going to ruin my business with other customers, I offered him the best terms I could. I was even willing to sell him parts at cost to try and patch things up. I mean… if I lost the business, I don't know what I would do. But he can be so stubborn. So I called him Saturday morning. I just thought it was time to work this out, one on one.

Lt. M.: When did you receive the note?

M. M.: A week ago Monday, I think.

Lt. M.: I see. Okay, Mr. Malone, that's all for now.

The modifications to the story did appear to have the desired effect, in that there was significant discussion regarding the solidity of Mickey's alibi. However, this did not seem to reduce the overall bias away from Mickey in the participants' pre-discussion opinions.

2 Experiment Instructions

Following the consent form, users were led through several pages of instructions, and following the experiment, an exit survey was provided. Each of these pages are included here for reference. In each of the following sections, I have provided pages from both adaptive and non-adaptive conditions if there were differences between the two conditions.

2.1 Introduction to the Training Application

Read Me!

Right click (Ctrl-click on macs) on this link to bookmark the login page now! If you have any problems at any point, return to the login page using this bookmark.

On the next page you will be given a decision problem for training purposes. The problem is to decide who would make the best president. The options are Paris Hilton, Ozzy Osbourne, and Dennis Rodman.

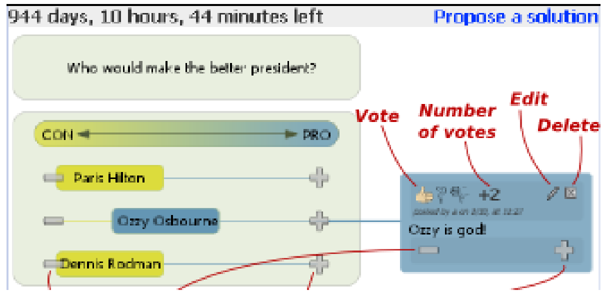
There will be instructions on the training page that guide you in the use of the software. Please read them carefully.

When everyone agrees on a decision option you will move on to the murder mystery. If you want to keep using the application, do not agree when consensus is proposed!

Take your time! You've a half an hour to play with application.

Figure 50: Text introducing the training application. There were no differences between the conditions

2.2 The Training Application



The box in the middle of the screen contains **answers** you must choose from as a team. Use the software to argue for and against each of the answers. **Arguments** appear as boxes that are connected either to answers or other arguments. Arguments that are blue are "for" arguments, arguments that are yellow are "against" arguments. The mystery is finished when you and your partners can agree on a solution.

Navigation

- Moving - Drag the mouse with the button held down while the cursor is on the white background.
- Zooming - Use the slider on the right to zoom in an out.
- Centering - Press the button above the slider
- Focusing- Clicking any node will highlight that node and make the other nodes it is attached to easier to see. Clicking the node again will make things normal again.

Arguing

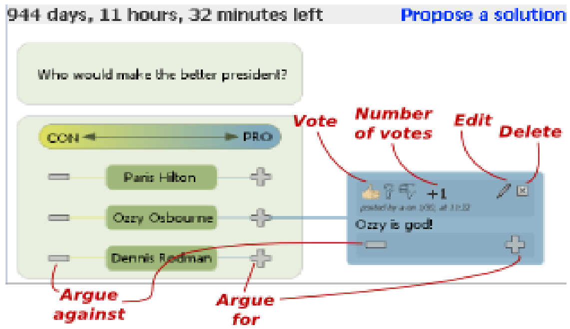
Use the system to argue about the possible answers. You can vote to change how the system evaluates arguments. You can reply to existing arguments to change the system evaluation and add information. The final assessment is determined by the system.

- Argue for or against an **answer** by clicking the plus(+) or minus(-) icon on either side of the answer
- Argue for or against an **argument** by clicking the plus(+) or minus(-) icon in the lower corners of the argument.
- To reiterate, click the "+" if you agree, and the "-" if you disagree with an existing post.
- Vote using "thumbs up" or "thumbs down" to change which answer is winning. The "question mark" means you have no opinion.
- Click "propose a solution" to propose a solution. **Everyone must agree with your proposal to successfully complete the problem.**

Important Tips

- Post short, simple statements. Do not use compound sentences with "and" "or" or "if."
- Post new information at the top of the thread.
- Only post arguments about an answer in the thread attached to that answer (i.e. do not talk about Ozzy Osbourne in Paris Hilton's thread).
- Reply to an argument you agree with by clicking "+." Reply to an argument you disagree with by clicking "-."
- The system's assessment is determined by your arguments and votes. The final solution you propose must match the system's assessment.

Figure 51: Text from Adaptive platform training application



The box in the middle of the screen contains **answers** you must choose from as a team. Use the software to argue for and against each of the answers. **Arguments** appear as boxes that are connected either to answers or other arguments. Arguments that are blue are "for" arguments, arguments that are yellow are "against" arguments. The mystery is finished when you and your partners can agree on a solution.

Navigation

- Moving - Drag the mouse with the button held down while the cursor is on the white background.
- Zooming - Use the slider on the right to zoom in an out.
- Centering - Press the button above the slider
- Focusing- Clicking any node will highlight that node and make the other nodes it is attached to easier to see. Clicking the node again will make things normal again.

Arguing

Use the system to argue about the possible answers. You can vote to change how the system evaluates arguments. You can reply to existing arguments to agree or disagree with information. At the end, you choose a final answer based on the information you exchange with others.

- Argue for or against an **answer** by clicking the plus(+) or minus(-) icon on either side of the answer
- Argue for or against an **argument** by clicking the plus(+) or minus(-) icon in the lower corners of the argument.
- Vote using "thumbs up" or "thumbs down" to express your agreement (or disagreement) with an argument. The "question mark" means you have no opinion.
- Click "propose a solution" to propose a solution. **Everyone must agree with your proposal to successfully complete the problem.**

Important Tips

- Post short, simple statements. Do not use compound sentences with "and" "or" or "if."
- Post new information at the top of the thread.
- Only post arguments about an answer in the thread attached to that answer (i.e. do not talk about Ozzy Osbourne in Paris Hilton's thread).
- Reply to an argument you agree with by clicking "+." Reply to an argument you disagree with by clicking "-."
- Try to assess all of the information when making a final proposal.

Figure 52: Text from Non-adaptive platform training application

2.3 Introduction to the Pre-assessment Task

Read Me!

On the next page you will read some information about a (fictional) murder. You may have some information other people don't have.

Some pieces of evidence implicate some suspects, and some pieces of evidence exonerate some suspects.

When you have read through your materials, you will choose the suspect you think "did it."

Once everyone has finished, you will move onto the collaborative portion, where you will try to decide on the most likely suspect.

Figure 53: Text introducing the pre-assessment task; there were no differences between the conditions.

2.4 Pre-assessment Instructions

Read Me!

Please familiarize yourself with the following set of evidence. After you have read through the available information, please indicate who you think the guilty suspect is.

You have a limited amount of time to complete this task (reported at the top of the screen). When you run out of time, the evidence will be removed and you should make your selection.

You **do not** need to memorize this information. It will be provided for you in the next part of the task.

Figure 54: Instructions provided during the pre-assessment task

2.5 Introduction to the Collaborative Task

Read Me!

- Post any new information at the top level! Do not introduce new clues at the bottom of a thread.
- Reply to an argument using "+" if you agree with it. Reply to an argument using "-" if you don't.
- Post simple messages. Don't use the words "and," "or," or "if."
- Do not talk about answer "A" in the thread attached to answer "B."
- If you post all the clues and follow these guidelines, the system's assessment will be correct.
- **You have until** <*timestamp*> **to complete this portion of the task.**

Figure 55: Instructions to groups in the Adaptive condition

Read Me!

- Post any new information at the top level! Do not introduce new clues at the bottom of a thread.
- Reply to an argument using "+" if you agree with it. Reply to an argument using "-" if you don't.
- Post simple messages. Don't use the words "and," "or," or "if."
- Do not talk about answer "A" in the thread attached to answer "B."
- **You have until** <*timestamp*> **to complete this portion of the task.**

Figure 56: Instructions to groups in the Non-adaptive condition

2.6 Exit Interview

Read Me!

1. Who do you think is guilty (even if your answer is different from the group's) ?
2. To what extent did the information contributed by others cause you to re-evaluate your choice (even if you did not change it)?
3. To what extent did the final consensus accurately reflect the combined information of the group (select "not at all" if consensus was not acheived)?
4. How familiar are you with mysteries?
5. How hard / easy was it to use the software?
6. How hard / easy was it to identify important information about the case?
7. How much did the system help you keep track of information about the case?
8. How much did the system help you keep track of the group's overall belief?
9. How hard / easy was it to combine everyone's information into a final assessment?
10. How hard / easy was it to achieve consensus?
11. Any additional comments (optional)?

In the exit interview, question one was answered via selection from a drop down list, and question eleven via a free-text field. Each of the remaining questions were by selection from a five point scale.

Appendix B

Chapter 7 introduced the notion of "focused agreement" areas in conversations as those areas where most of the conversation was confined to a single physical area in the conversation, there was a substantial amount of agreement, and posts occurred rapidly. The claim was also made that focused agreement areas corresponded to regions where some type of "critical story construction" took place for the group. Groups 2 and 15 were used as examples, and Group 22 was offered as an exception that did not appear to fit the pattern.

In this following sections, I examine the peaks of focused agreement and were not covered in Chapter 7. Groups 20, 19, 11, and 13 are examined.

1 Group 20

Post #	Author	Valence	Post
1	BE	[top – implicates B]	they found a crowbar with his fingerprints on it.
71	BE	CON	too obvious. they would never give us a silver bullet. instead, thats just misdirection.
77	S1	PRO	yes
95	BE	PRO	Also, motive?
97	EE	CON	Occam's razor [*deleted by BE*]
98	ME	PRO	Finally, crowbar can be shared by laborers – maybe lots of people's prints are on it... we have no negation of that possibility...
99	ME	PRO	needed money for the ponies
101	BE	CON	don't delete my posts.. dick.
102	S2	CON	DUDE has a gambling problem he say he doesnt, but ask a drunk if they think they are an alcoholic!!!
106	BE	PRO	who claims he has a gambling problem? and worth it to murder, as opposed to simply steal? and for 50 bucks?
107	EE	CON	Occam's razor is still an argument against it, douche
108	ME	PRO	He did. The man was in debt. he was always asking for advances.
109	BE	PRO	Never said is wasn't... just need to hear HOW .. make sure we're on the same page
110		PRO	I agree!!!!!!!!! timmy
111	BE	CON	did he say that himself in his affadavit? if so, than i'll consider him, but still, worth it to murder?
112	S2	CON	yeah dude needs money it rained the night before at the track likely... course may have been sloppy...and long shots coming in like relatives
113		PRO	yes he admitted to asking for advances and the wife confirmed
--thread continues--			

Table 41: Group 20 focus area

Group 20 did not do a very good job at weaving together a story. The group was comprised of five friends who had a very playful style of interaction, and did not seem to be focused on the task. During the pre-study period, the BE participant refused to give his clues more than a cursory overview, despite the requests of the experimenter to concentrate on the task. There were also several instances where one participant deleted another participant's posts (noted in Table 41), counter to the experimenter's requests. This may be one of the drawbacks of studying group decision making in a university environment.

There is very little in the way of story creation for Group 20, and although all of the unshared information about E is mentioned at some point, most of this gets ignored. Nonetheless, the area of focused agreement noted in Table 41 seems to be the primary justification for the group's final answer, which is B, because it is the simplest explanation, and he clearly had a motive.

2 Group 19

Post #	Time	Author	Valence	Post
20	17:19:12	ME	[top – implicates M]	No one can corroborate where he was between leaving his house early saturday and arriving at the country club later
45	17:31:37	S1	PRO	It says he got all the way to Crestview before turning around. That is pretty close to Blake's house.
49	17:35:38	ME	CON	I have a golf partner saying that he (the partner arrived at 7am and that Mickey had already been there at least 10 minutes).
50	17:36:45	BE	PRO	Yea, so what? The murder took place between 6:30 and 7 ...
51	17:37:00	EE	PRO	I don't have anything about that in mine.
53	17:38:18	EE	PRO	was there time for him to kill the guy before he went to golf?
54	17:38:27	ME	PRO	Right, and Billy Prince says he found the body at around 8am, and Eddie and Mrs. Blake at about 10:30.
55	17:38:33	BE	CON	Mickey admits to coming early
56	17:38:56	ME	CON	And drive by the shop dumpster to drop off the wallet.
57	17:39:32	S2	CON	why would he take the wallet? Only Billy has any incentive to take it.
58	17:39:34	BE	PRO	Yea. Mickey on phone at 6, then left house ... but we need to explain Mary Blake's observations in negative.
62	17:41:29	S1	PRO	Mary Blake heard the car drive away at 6:40, so the murder [happened] before that
65	17:41:58	ME	CON	Or right at 6:40
66	17:42:37	S1	PRO	Mickey could have left the house at 6:40 and still gotten to the golf course at 6:50
--thread continues--				

Table 42: Group 19 focus area

Group 19 chooses M, based upon their inability to establish when M left his house to drive to Gill's, and reasoning that he had an obvious motive and the clearest opportunity.

The focused agreement area beginning at post 53 is the point in the conversation where the group appears to form consensus around the notion that Mickey does not have an air-tight alibi, and hence had time to commit the murder.

3 Group 11

Time	Author	Valence	Post
14:13:19	S2	[top level implicates M]	Was already near Blake's house – could've killed him and still arrived at the golf course on time.
14:14:32	ME	PRO	true
14:15:37	BE	OON	But the crowbar had to be thrown in the bushes after billy arrived and got his prints on it.
14:16:36	ME	CON	so billy came after mm and threw the crowbar
14:18:53	S1	CON	But did he?
14:19:40	S2	PRO	why would billy throw the crowbar?
14:19:48	ME	PRO	billy was there on sat. morning, and he was there after mm would have gotten there.
14:20:30	ME	PRO	because he tried to get it out of the way and he got scared/upset
19:20:45	EE	CON	Billy was also there on Friday right? MM might have known this an thrown the crowbar with Billy's fingerprints on it.
The following post appears in a different thread at this point.			
19:21:32	EE	[top exonerates E]	He would not throw the crowbar into the bushes if he had done it
The following continues the original thread.			
19:21:49	S1	CON	Why would a crowbar be “in the way”? It doesn't take up much space.
14:22:26	EE	CON	I think MM threw the crowbar
14.23:21	S2	PRO	It would make sense if we understood why billy's fingerprints were on it. Billy said that ES always keeps his tools locked up, so why would there be prints?

Table 43: Focus thread for Group 11; rapid posting area is highlighted.

Group 11 presents an interesting case. The region of rapid posting in the focus thread cooccurs with a split into sub-threads, as participants try to work out an inconsistency in their story. The inconsistency is not resolved, but becomes part of the final answer regardless. This focus area immediately precedes a period of explicit conversation about the participants' respective beliefs, and the team ultimately chooses M as the answer.

4 Group 13

In comparison to the other groups, Group 13 is highly collaborative and performs an exemplary analysis of the murder. The collaborators determine that E framed B, both with the crowbar and the muffler. They determine that the wallet is M's alibi, but never quite get to the point of realizing that it is part of E's attempt to frame B. The conversation has a very clear focus thread, and a follow-up post that summarizes their story.

One of the first several posts in the conversation, prior to the focus thread, identified the connection between E claiming to have heard B's muffler, and his hearing problem. Thirteen minutes after this inference was made, another participant brought up the possibility that E framed B, but that suggestion did not receive a lot of attention. The clue about the crowbar (B said he touched the crowbar) was also mentioned early on, but received no responses.

Post	Time	Author	Valence	Post
70	19:20:30	S2	[top level implicates E]	very much disregarded the question did he find the crowbar
73	19:21:34	S1	CON	how would billie's fingerprints get on eddie's crowbar?
75	19:22:30	BE	CON	Billy said he moved the crowbar to get to the mower.
76	19:23:06	S1	CON	and would that really require throwing the murderweapo- crowbar into bushes, to be hidden?
77	19:23:51	EE	CON	no, but eddie might of used the opportunity to frame him
78	19:24:10	BE	CON	He never said he threw it into the bushes, just that he moved it - Eddit could move it
79	19:24:48	BE	PRO	Eddie also wants to pin this on Billy from the beginning. (muffler)
80	19:25:16	S1	CON	what's eddie's motivation to frame billy?
82	19:26:04	BE	CON	It's somebody that is not Eddie.
83	19:2632	S1	PRO	lolpwnd
84	19:27:06	S2	CON	easy scapegoat

Table 44: Focus thread for Group 13.

The focus thread shown in Table 44 is the second to the last thread that occurs. At this point in the conversation, the group has figured out that M has an alibi, and that E may have a motive regarding his daughter. The thread begins with an observation that E seemed fairly nonchalant about his missing crowbar, and uses this as an argument for E as the culprit. S1 rebuts with a question that had been previously brought up by the group – how B's fingerprints got on the crowbar. BE replies with the unshared clue about B moving the crowbar. Three posts later, this is connected to the E's attempt to frame B by suggesting that he had heard B's muffler that morning. S1 asks about E's motivation to frame B, to which BE and S2 reply that framing shifts attention away from E. S1 follows up with lighthearted "lolpwnd[7]," bringing closure to the conversation.

7 "Lol" - laugh out loud; Pwnd - Owned; beaten soundly. My interpretation is that S1 is admitting to the others it was kind of a silly question.

References

1. Albrecht, David W., Ingrid Zukerman, and An E. Nicholson. 1998. Bayesian Models for Keyhole Plan Recognition in an Adventure Game. *User Modeling and User-Adapted Interaction* 8, no. 1 (March 1): 5-47. doi:10.1023/A:1008238218679.
2. Alterman, R. 2000. Rethinking Autonomy. *Minds and Machines* 10, no. 1: 15-30.
3. Alterman, R., A. Feinman, J. Introne, and S. Landsman. 2001. Coordinating Representations in Computer-Mediated Joint Activities. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, 43-48.
4. Anderson, John R., Albert T. Corbett, Kenneth R. Koedinger, and Ray. Pelletier. 1995. Cognitive Tutors: Lessons Learned. *Journal of the Learning Sciences* 4, no. 2: 167. doi:10.1207/s15327809jls0402_2.
5. Baker, M., and K. Lund. 1997. Promoting reflective interactions in a CSCL environment. *Journal of Computer Assisted Learning* 13, no. 3: 175-193. doi:10.1046/j.1365-2729.1997.00019.x.
6. Bales, R. F., and F. L. Strodtbeck. 1951. Phases in group problem-solving. *J Abnorm Soc Psychol* 46, no. 4: 485-95.
7. Bass, L., R. Pellegrino, S. Reed, et al. 1991. The Arch Model: Seeheim Revisited. *User Interface Developers' Workshop*.
8. Bauer, E., D. Koller, and Y. Singer. 1997. Update Rules for parameter estimation in Bayesian networks. In *Proceedings of the 13th Annual Conference on Uncertainty in AI (UAI)*, 3-13. http://citeseer.ist.psu.edu/bauer97update.html.
9. Bell, D. E., H. Raiffa, and A. Tversky. 1988. Descriptive, Normative, and Prescriptive Interactions in Decision Making. In *Decision Making: Descriptive, Normative, and Prescriptive Interactions*. Cambridge University Press.
10. Benyon, David. 1993. Adaptive systems: A solution to usability problems. *User Modeling and User-Adapted Interaction* 3, no. 1 (March 1): 65-87. doi:10.1007/BF01099425.
11. Billsus, Daniel, and Michael J. Pazzani. 2000. User Modeling for Adaptive News

Access. *User Modeling and User-Adapted Interaction* 10, no. 2 (June 1): 147-180. doi:10.1023/A:1026501525781.

12. Bosma, Wauter, and Elisabeth André. 2004. Exploiting emotions to disambiguate dialogue acts. In *Proceedings of the 9th international conference on Intelligent user interfaces*, 85-92. Funchal, Madeira, Portugal: ACM. doi:10.1145/964442.964459. http://portal.acm.org/citation.cfm?id=964459.
13. Bowman, J. M., and G. M. Wittenbaum. 2002. Effects of time pressure and informational access on hidden profile solutions in decision-making groups. In *Annual meeting of the National Communication Association*. New Orleans, LA.
14. Bødker, S. 1989. A Human Activity Approach to User Interfaces. *Human-Computer Interaction* 4, no. 3: 171-195.
15. ---. 1990. *Through the Interface: A Human Activity Approach To User Interface Design.* Lawrence Erlbaum Associates.
16. Brehmer, B., and C. R. B. Joyce. 1988. *Human judgment: the SJT view.* Amsterdam: North-Holland.
17. Brennan, S. E. 1998. The Grounding Problem in Conversation Wth and Through Computers. In , ed. Susan R. Fussell and Roger J. Kreuz, 201-225. Hillsdale, NJ: Lawrence ErlbaumAssociates.
18. Briggs, R. O. 1994. The focus theory of group productivity and its application to development and testing of electronic group support systems. University of Arizona.
19. ---. 2006. On theory-driven design and deployment of collaboration systems. *International Journal of Human-Computer Studies* 64, no. 7: 573-582.
20. Brodbeck, F. C., R. Kerschreiter, A. Mojzisch, D. Frey, and S. Schulz-Hardt. 2002. The dissemination of critical, unshared information in decision-making groups: the effects of pre-discussion dissent. *European Journal of Social Psychology* 32, no. 1: 35-56.
21. Browne, D., M. Norman, D. Riches, et al. 1990. Why Build Adaptive Systems. In *Adaptive user interfaces*, 15-58.
22. Buckingham Shum, S. J. 2003. The Roots of Computer Supported Argument Visualization. In *Visualizing argumentation: software tools for collaborative and educational sense-making*, ed. P. A. Kirschner, S. J. Buckingham Shum, and C. S. Carr, 3-24. Springer-Verlag. http://portal.acm.org/citation.cfm?id=779336.
23. Buckingham Shum, S. J., A. Selvin, M. Sierhuis, et al. 2006. Hypermedia Support for Argumentation-Based Rationale. In *Rationale Management in Software Engineering*, ed. A. Dutoit, R. McCall, I. Mistrik, and B. Paech, 111-132. http://dx.doi.org/10.1007/978-3-540-30998-7_5.
24. Buede, D. M., and D. T. Maxwell. 1995. Rank disagreement: A comparison of multi-criteria methodologies. *Journal of Multi-Criteria Decision Analysis*

4, no. 1: 1-21.

25. Burnstein, E., and A. Vinokur. 1973. Testing two classes of theories about group induced shifts in individual choice. *Journal of Experimental Social Psychology* 9, no. 2: 123-137.
26. Carr, C. S. 2003. Using computer supported argument visualization to teach legal argumentation. In *Visualizing argumentation: software tools for collaborative and educational sense-making*, ed. P. A. Kirschner, S. J. Buckingham Shum, and C. S. Carr, 75-96. Springer-Verlag. http://portal.acm.org/citation.cfm?id=779336.
27. Carroll, J. M., D. C. Neale, P. L. Isenhour, M. B. Rosson, and D. S. McCrickard. 2003. Notification and awareness: synchronizing task-oriented collaborative activity. *International Journal of Human-Computer Studies* 58, no. 5: 605-632.
28. Cesta, Amedeo, and Daniela D'Aloisi. 1999. Mixed-Initiative Issues in an Agent-Based Meeting Scheduler. *User Modeling and User-Adapted Interaction* 9, no. 1 (April 1): 45-78. doi:10.1023/A:1008359001692.
29. Chalmers, M. 2002. Awareness, Representation and Interpretation. *Computer Supported Cooperative Work (CSCW)* 11, no. 3: 389-409. doi:10.1023/A:1021209028381.
30. Charniak, Eugene, and Robert P. Goldman. 1993. A Bayesian model of plan recognition. *Artif. Intell.* 64, no. 1: 53-79.
31. Chin, D. N. 1986. User modeling in UC, the UNIX consultant. *SIGCHI Bull.* 17, no. 4: 24-28. doi:10.1145/22339.22343.
32. Clark, H. H. 1996. *Using Language*. Cambridge University Press.
33. Clark, H. H., and S. A. Brennan. 1991. Grounding in communication. In *Perspectives on Socially Shared Cognition*, ed. L. Resnick, J. Levine, and S. Teasley, 127-149. APA.
34. Clark, H. H., and D. Wilkes-Gibbs. 1990. Referring as a collaborative process. In *Intentions in Communication*, ed. P. R. Cohen, J. L. Morgan, and M. E. Pollack, 463-493. MA: MIT Press.
35. Cohen, M. S. 1993. Three paradigms for viewing decision biases. In *Decision making in action: Models and methods*, ed. G. A. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok, 36-50. Ablex Publishing.
36. Cohen, P. R., C. R. Perrault, and J. F. Allen. 1981. *Beyond Question-Answering.* Cambridge, MA: BBN Inc.
37. Cole, M., and Y. Engeström. 1993. A cultural-historical approach to distributed cognition. In *Distributed cognitions: Psychological and educational considerations*, ed. Gavriel Salomon, 1-46. Cambridge University Press.
38. Collins, A., and R. Michalski. 1989. The logic of plausible reasoning: A core theory. *Cognitive Science* 13, no. 1 (March): 1-49. doi:10.1016/0364-0213(89)90010-4.

39. Conati, Cristina, Abigail Gertner, and Kurt VanLehn. 2002. Using Bayesian Networks to Manage Uncertainty in Student Modeling. *User Modeling and User-Adapted Interaction* 12, no. 4 (November 1): 371-417. doi:10.1023/A:1021258506583.
40. Conklin, J. 2003. Dialog Mapping: Reflections on an Industrial Strength Case Study. In *Visualizing argumentation: software tools for collaborative and educational sense-making*, ed. P. A. Kirschner, S. J. Buckingham Shum, and C. S. Carr, 117-136. Springer-Verlag. http://portal.acm.org/citation.cfm?id=779336.
41. Conklin, J., and M. L. Begeman. 1988. gIBIS: a hypertext tool for exploratory policy discussion. *ACM Transactions on Information Systems (TOIS)* 6, no. 4: 303-331.
42. Cummings, M. L. 2004. The Need for Command and Control Instant Message Adaptive Interfaces: Lessons Learned from Tactical Tomahawk Human-in-the-Loop Simulations. *CyberPsychology & Behavior* 7, no. 6: 653-661.
43. Cummings, M. L., and P.J. Mitchell. 2007. Operator scheduling strategies in supervisory control of multiple UAVs. *Aerospace Science and Technology* 11, no. 4 (May): 339-348. doi:10.1016/j.ast.2006.10.007.
44. Cypher, A. 1991. EAGER: programming repetitive tasks by example. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*, 33-39. New Orleans, Louisiana, United States: ACM. doi:10.1145/108844.108850. http://portal.acm.org/citation.cfm?id=108850.
45. Das, S. 2005. Symbolic Argumentation for Decision Making under Uncertainty. In *7th International Conference on Information Fusion, 2005* , 2:
46. Das, S. K., J. Fox, D. Elsdon, and P. Hammond. 1997. A flexible architecture for autonomous agents. *Journal of Experimental & Theoretical Artificial Intelligence* 9, no. 4: 407-440.
47. Davis, D. N., and S. C. Lewis. 2003. Computational Models of Emotion for Autonomy and Reasoning. *Informatica* 27, no. 2: 159-165.
48. Davis, J. H. 1973. Group decision and social interaction: A theory of social decision schemes. *Psychological Review* 80, no. 2: 97-125.
49. Dennis, A. R. 1996. Information Exchange and Use in Group Decision Making: You Can Lead a Group to Information, but You Can't Make It Think. *MIS Quarterly* 20, no. 4 (December): 433-457.
50. Dennis, A. R., K. M. Hilmer, and N. J. Taylor. 1997. Information exchange and use in GSS and verbal group decision making: effects of minority influence. *Journal of Management Information Systems* 14, no. 3: 61-88.
51. DeSanctis, G., and R. B. Gallupe. 1987. A foundation for the study of group decision support systems. *Management Science* 33, no. 5: 589-609.
52. Deshpande, N., B. de Vries, and J. P. van Leeuwen. 2005. Building and Supporting Shared Understanding in Collaborative Problem-Solving.

Information Visualisation, 2005. Proceedings. Ninth International Conference on: 737-742.

53. Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77, no. 2: 321-357.
54. Edmonds, E. A. 1992. *The separable user interface*. London: Academic Press.
55. Einhorn, H. J., and R. M. Hogarth. 1978. Confidence in Judgment: Persistence of the Illusion of Validity. *Psychological Review* 85, no. 5: 395-416.
56. Ellis, C., S. Gibbs, and G. Rein. 1991. Groupware: Some Issues and Experiences. *Communicationsof the ACM* 34, no. 1: 38-58.
57. Ellis, C., S. J. Gibbs, and G. Rein. 1990. Design and use of a group editor. *Engineering for Human-Computer Interaction: Proceedings of the IFIP TC 2/WG 2.7 Working Conference on Engineering for Human-Computer Interaction, Napa Valley, California, USA, 21-25 August, 1989*.
58. Engeström, Yrjo. 2000. Activity theory as a framework for analyzing and redesigning work. *Ergonomics* 43, no. 7: 960. doi:10.1080/001401300409143.
59. Engeström, Yrjo, and David Middleton. 1996. *Cognition and Communication at Work: studying work as mindful practice*. Cambridge UniversityPress.
60. Evans, E. 2004. *Domain-Driven Design: Tackling Complexity in the Heart of Software*. Addison-Wesley Professional.
61. Feinman, A. 2006. From Discourse Analysis to Groupware Design. Brandeis University.
62. Feinman, A., and R. Alterman. 2003. Discourse Analysis Techniques for Modeling Group Interaction. In *User Modeling 2003*, 147. http://dx.doi.org/10.1007/3-540-44963-9_31.
63. Ferguson, George, and James F. Allen. 1998. TRIPs: an integrated intelligent problem-solving assistant. In *Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence*, 567-572. Madison, Wisconsin, United States: American Association for Artificial Intelligence. http://portal.acm.org/citation.cfm?id=295737&dl=.
64. Ferguson, George, James F. Allen, and Brad Miller. 1996. TRAINS-95: Towards a Mixed-Initiative Planning Assistant. In *Artificial Intelligence Planning Systems*, 70-77. http://citeseer.ist.psu.edu/ferguson96train.html.
65. Festinger, Leon. 1975. *A theory of social comparison processes*. [Indianapolis, Ind.]: Bobbs-Merrill.
66. Fink, Josef, Alfred Kobsa, and Andreas Nill. 1988. Adaptable and adaptive information provision for all users, including disabled and elderly people. *New Review of Hypermedia and Multimedia* 4, no. 1: 163-188. doi:10.1080/13614569808914700.

67. Fischer, Gerhard. 2001. User Modeling in Human–Computer Interaction. *User Modeling and User-Adapted Interaction* 11, no. 1 (March 1): 65-86. doi:10.1023/A:1011145532042.

68. Fjermestad, J. 2004. An analysis of communication mode in group support systems research. *Decision Support Systems* 37, no. 2: 239-263.

69. Fjermestad, J., and S. R. Hiltz. 1998. An assessment of group support systems experimental research: methodology and results. *Journal of Management Information Systems* 15, no. 3: 7-149.

70. Flores, F., M. Graves, B. Hartfield, and T. Winograd. 1988. Computer systems and the design of organizational interaction. *ACM Trans. Inf. Syst.* 6, no. 2: 153-172. doi:10.1145/45941.45943.

71. Fraidin, Samuel N. 2004. When is one head better than two? Interdependent information in group decision making*1. *Organizational Behavior and Human Decision Processes* 93, no. 2 (March): 102-113. doi:10.1016/j.obhdp.2003.12.003.

72. Ganoe, C. H., J. P. Somervell, D. C. Neale, et al. 2003. Classroom BRIDGE: using collaborative public and desktop timelines to support activity awareness. *Proceedings of the 16th annual ACM symposium on User interface software and technology*: 21-30.

73. van Gelder, T. 2003. Enhancing Deliberation Through Computer Supported Visualization. In *Visualizing argumentation: software tools for collaborative and educational sense-making*, ed. P. A. Kirschner, S. J. Buckingham Shum, and C. S. Carr, 97-115. Springer-Verlag. http://portal.acm.org/citation.cfm?id=779336.

74. Gigone, D., and R. Hastie. 1993. The common knowledge effect: Information sharing and group judgment. *Journal of Personality and Social Psychology* 65, no. 5: 959-974.

75. ---. 1997. The impact of information on small group choice. *Journal of personality and social psychology* 72, no. 1: 132-140.

76. Goodman, B. A., F. N. Linton, R. D. Gaimari, et al. 2005. Using Dialogue Features to Predict Trouble During Collaborative Learning. *User Modeling and User-Adapted Interaction* 15, no. 1: 85-134.

77. Goodwin, C., and M. H. Goodwin. 1996. Seeing as situated activity: Formulating planes. In *Cognition and communication at work*, ed. Yrjo Engestrom and David Middleton, 61-95. Cambridge University Press.

78. Green, Mark. 1986. A survey of three dialogue models. *ACM Trans. Graph.* 5, no. 3: 244-275. doi:10.1145/24054.24057.

79. Greitemeyer, T., and S. Schulz-Hardt. 2003. Preference-consistent evaluation of information in the hidden profile paradigm: Beyond group-level explanations for the dominance of shared information in group decisions. *Journal of Personality and Social Psychology* 84, no. 2: 322-339.

80. Hackman, J. R., and R. E. Kaplan. 1974. Interventions into Group Process: An

Approach to Improving the Efectiveness of Groups. *Decision Sciences* 5, no. 3: 459-480.

81. Hammond, K. R. 1988. Judgment and decision making in dynamic tasks. *Information and decision technologies* 14, no. 1: 3-14.
82. ---. 1993. Naturalistic decision making from a Brunswikian viewpoint: Its past, present, future. In *Decision making in action: Models and methods*, ed. G. A. Klein, Judith Orasanu, Roberta Calderwood, and C. E. Zsambok, 205-227.
83. Hastie, R., and N. Pennington. 2000. Explanation-based decision making. *Judgment and decision making: An interdisciplinary reader* 2: 212-228.
84. Heacox, N. J., R. A. Moore, J. G. Morrison, and R. F. Yturralde. 2004. Real-time Online Communications: "Chat" Use in Navy Operations. In *Proceedings of Command and Control Research and Technology Symposium*. San Diego, CA. . http://www. dodccrp. org/events/2004/CCRTS_San_Diego/CD/papers/086. pdf.
85. Hefley, William E., and Dianne Murray. 1993. Intelligent user interfaces. In *Proceedings of the 1st international conference on Intelligent user interfaces*, 3-10. Orlando, Florida, United States: ACM. doi:10.1145/169891.169892. http://portal.acm.org/citation.cfm?id=169892.
86. Heidegger, M. 1962. *Being and Time*. Trans. J. Macquarrie and E. Robinson. New York: Harper & Row.
87. Hightower, R., and L. Sayeed. 1996. Effects of communication mode and prediscussion information distribution characteristics on information exchange in groups. *Information Systems Research* 7, no. 4: 451-465.
88. Hirokawa, R. Y., and M. S. Poole. 1986. *Communication and Group Decision-Making*. Sage Publications.
89. Hogarth, R. M., and H. J. Einhorn. 1992. Order Effects in Belief Updating: The Belief-Adjustment Model. *Cognitive Psychology* 24, no. 1: 1-55.
90. Hollingshead, A. 1996. The Rank-Order Effect in Group Decision Making. *Organizational Behavior and Human Decision Processes* 68, no. 3 (December): 181-193. doi:10.1006/obhd.1996.0098.
91. Horvitz, E. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, 159-166. Pittsburgh, Pennsylvania, United States: ACM. doi:10.1145/302979.303030. http://portal.acm.org/citation.cfm?id=303030.
92. Horvitz, E., J. Breese, D. Heckerman, D. Hovel, and K. Rommelse. 1998. The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, 256—265. Madison, WI, July. http://citeseer.ist.psu.edu/horvitz98lumiegravere.html.

93. Hutchins, E. 1995a. How a cockpit remembers its speeds. *Cognitive Science* 19, no. 3: 265-288.

94. ---. 1995b. *Cognition in the Wild*. Bradford Books.

95. Hutchins, E., J. D. Hollan, and D. A. Norman. 1985. Direct Manipulation Interfaces. *Human-Computer Interaction* 1, no. 4: 311. doi:10.1207/s15327051hci0104_2.

96. Introne, J., and R. Alterman. 2004. Leveraging a better interface language to simplify adaptation. *Proceedings of the 9th international conference on Intelligent user interface*: 262-264.

97. ---. 2006. Using shared representations to improve coordination and intent inference. *User Modeling and User-Adapted Interaction* 16, no. 3: 249-280. doi:10.1007/s11257-006-9009-2.

98. Jameson, A. 2003. Adaptive interfaces and agents. In *Human-computer interaction handbook*, ed. JA Jacko and A Sears, 305-330.

99. Janis, I. L. 1982. *Groupthink*. Houghton Mifflin Boston.

100. Jones, P. M., C. C. Hayes, C. Wilkins, et al. 1998. CoRAVEN: modeling and design of a multimedia intelligent infrastructure for collaborative intelligence analysis. In *1998 IEEE International Conference on Systems, Man, and Cybernetics, 1998.*, 1:

101. Kahneman, D., and A. Tversky. 1982. The psychology of preferences. *Scientific American* 246, no. 1: 160-173.

102. Kaiser, Edward C. 2005. Multimodal new vocabulary recognition through speech and handwriting in a whiteboard scheduling application. In *Proceedings of the 10th international conference on Intelligent user interfaces*, 51-58. San Diego, California, USA: ACM. doi:10.1145/1040830.1040851. http://portal.acm.org/citation.cfm?id=1040830.1040851.

103. Kerr, N. L., and R. S. Tindale. 2004. Group Performance and Decision Making. Review-article. January 12. http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.psych.55.090902.142009?cookieSet=1&journalCode=psych.

104. Klein, G. A. 1993. A recognition-primed decision (RPD) model of rapid decision making. In *Decision making in action: Models and methods*, ed. G. A. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok, 138-147. Ablex Publishing.

105. Klein, G. A., J. Orasanu, and R. Calderwood. 1993. *Decision Making in Action: Models and Methods*. Ablex Publishing, January 1.

106. Krasner, G., and S. Pope. 1988. A Description of the Model-View-Controller User Interface Paradigm in the Smalltalk-80 system. *Journal of Object Oriented Programming* 1, no. 3: 26—49.

107. Kuhn, D. 1991. *The Skills of Argument*. Cambridge University Press.

108. Kunz, W., and H. Rittel. 1970. *Issues as elements of information systems*. Center for Planning and Development Research, University of California at Berkeley.

109. Lagroue, H. J. 2006. The Effectiveness of Virtual Facilitation in Supporting GDSS Appropriation and Structured Group Decision Making. Louisiana State University.

110. Lam, S. S., and J. Schaubroeck. 2000. Improving group decisions by better pooling information: a comparative advantage of group decision support systems. *J Appl Psychol* 85, no. 4: 565-73.

111. Lamm, H., and G. Trommsdorff. 1973. Group versus individual performance on tasks requiring ideational proficiency (brainstorming): A review. *European Journal of Social Psychology* 3, no. 4: 361-388.

112. Landsman, S., and R. Alterman. 2005. *Using Transcription and Replay in Analysis of Groupware Applications*. Brandeis University.

113. Landsman, S., R. Alterman, A. Feinman, and J. Introne. 2000. *Vesselworld and ADAPTIVE (2000)*. Demo given at CSCW 2000. Brandeis University .

114. Langley, P. 1999. User modeling in adaptive interfaces. *Proceedings of the Seventh International Conference on User Modeling*: 357-370.

115. Larson Jr. , J. R. 1997. Modeling the Entry of Shared and Unshared information into Group Discussion: A Review and Basic Language Computer Program. *Small Group Research* 28, no. 3: 454-479.

116. Larson Jr. , J. R., C Christensen, A S Abbott, and T M Franz. 1996. Diagnosing groups: charting the flow of information in medical decision-making teams. *Journal of Personality and Social Psychology* 71, no. 2 (August): 315-30. doi:8765484.

117. Lau, Tessa, Steven A. Wolfman, Pedro Domingos, and Daniel S. Weld. 2001. Learning repetitive text-editing procedures with SMARTedit. In *Your wish is my command: programming by example*, ed. Henry Lieberman, 209-226. Morgan Kaufmann Publishers Inc. http://portal.acm.org/citation.cfm?id=369519.

118. Lee, Jintae. 1990. SIBYL: a tool for managing group design rationale. In *Proceedings of the 1990 ACM conference on Computer-supported cooperative work*, 79-92. Los Angeles, California, United States: ACM. doi:10.1145/99332.99344. http://portal.acm.org/citation.cfm?id=99344.

119. Lesh, N., O. Etzioni, L. C. Aiello, J. Doyle, and S. Shapiro. 1996. Scaling up Goal Recognition. *KR'96: Principles of Knowledge Representation and Reasoning*: 244-255.

120. Lesh, Neal, Charles Rich, and Candace L. Sidner. 1999. Using plan recognition in human-computer collaboration. In *Proceedings of the seventh international conference on User modeling*, 23-32. Banff, Canada: Springer-Verlag New York, Inc. http://portal.acm.org/citation.cfm?id=317331.

121. Limayem, M., and G. DeSanctis. 2000. Providing Decisional Guidance for Multicriteria Decision Making in Groups. *Info. Sys. Research* 11, no. 4: 386-401.

122. Lipshitz, R. 1993. Converging Themes in the Study of Decision Making in Realistic Settings. In *Decision making in action: Models and methods*, ed. G. A. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok, 103-137. Ablex Publishing.

123. Lipshitz, R., and M. S. Cohen. 2005. Warrants for Prescription: Analytically and Empirically Based Approaches to Improving Decision Making. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 47 (Spring): 102-120. doi:10.1518/0018720053653811.

124. Litman, Diane J., and Shimei Pan. 2000. Predicting and Adapting to Poor Speech Recognition in a Spoken Dialogue System. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, 722-728. AAAI Press / The MIT Press. http://portal.acm.org/citation.cfm?id=647288.721283.

125. Lord, C. G., L. Ross, and M. R. Lepper. 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* 37, no. 11: 2098-2109.

126. Lowrance, J., I. Harrison, A. Rodriguez, et al. 2008. Template-Based Structured Argumentation. In *Knowledge Cartography: Software Tools and Mapping Techniques*, ed. A. Okada, S. J. Buckingham Shum, and T. Sherborne. Advanced Information and Knowledge porcessing. Springer.

127. M. L. Cummings. 2005. The Need for Command and Control Instant Message Adaptive Interfaces: Lessons Learned from Tactical Tomahawk Human-in-the-Loop Simulations. Research-article. February 1. http://www.liebertonline.com/doi/abs/10.1089/cpb.2004.7.653.

128. Maass, A., and R. D. Clark. 1984. Hidden impact of minorities: 15 years of minority influence research. *Psychological Bulletin* 95, no. 3: 428-450.

129. Maes, P. 1994. Agents that reduce work and information overload. *Communications of the ACM* 37, no. 7: 30-40.

130. Maes, P., and B. Schneiderman. 1997. Direct Manipulation vs. Interface Agents: A Debate. *Interactions* 4, no. 6.

131. Malone, Thomas W. 2001. Two Design Principles for Collaboration Technology: Examples of Semiformal Systems and Radical Tailorability. In *Coordination Theory and Collaboration Technology*, ed. Gary M. Olson, Thomas W. Malone, and John B. Smith, 125-160. NJ: Lawrence Erlbaum Associates.

132. Masthoff, Judith. 2004. Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. *User Modeling and User-Adapted*

Interaction 14, no. 1 (February 1): 37-85. doi:10.1023/B:USER.0000010138.79319.fd.

133. Mennecke, B. E. 1997. Using group support systems to discover hidden profiles: an examination of the influence of group size and meeting structures on information sharing and decision quality. *International Journal of Human-Computer Studies* 47, no. 3: 387-405.
134. Miah, T., M. Karageorgou, and R. P. Knott. 1997. Adaptive Toolbars: An Architectural Overview. In *3rd ERCIM Workshop on User Interfaces for All*, 157-163.
135. Myers, D. G., and H. Lamm. 1976. The group polarization phenomenon. *Psychological Bulletin* 83, no. 4: 602-627.
136. Newman, S. E., and C. C. Marshall. 1991. *Pushing Toulmin Too Far: Learning From an Argument Representation Scheme*. Palo Alto, CA: Xerox PARC.
137. Noble, D. 1993. A Model to Support Development of Situation Assessment Aids. In *Decision making in action: Models and methods*, ed. G. A. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok, 287-305. Ablex Publishing.
138. Norman, D. A. 1976. *Memory and attention*. Wiley New York.
139. ---. 1991. Cognitive artifacts. In *Designing interaction: psychology at the human-computer interface*, ed. J. M. Carroll, 17-38. Cambridge University Press. http://portal.acm.org/citation.cfm?id=120352.120354.
140. Nunamaker, J. F., A. R. Dennis, J. S. Valacich, D. Vogel, and J. F. George. 1991. Electronic meeting systems. *Commun. ACM* 34, no. 7: 40-61. doi:10.1145/105783.105793.
141. Oppermann, R. 1994. Adaptively supported adaptability. *International Journal of Human-Computer Studies* 40, no. 3: 455-472.
142. Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers.
143. Pennington, N. 1981. Causal reasoning and decision making: the case of juror decisions. Harvard Graduate School of Education.
144. Pennington, N., and R. Hastie. 1986. Evidence evaluation in complex decision making. *Journal of personality and social psychology* 51, no. 2: 242-258.
145. ---. 1988. Explanation-based decision making: effects of memory structure on judgment. *Journal of experimental psychology. Learning, memory, and cognition* 14, no. 3: 521-533.
146. ---. 1992. Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology* 62, no. 2: 189-206.
147. ---. 1993. Reasoning in explanation-based decision making. *Cognition* 49, no. 1-2: 123-163.
148. Perkowitz, Mike, and Oren Etzioni. 2000. Adaptive Web sites. *Commun. ACM* 43, no. 8: 152-158. doi:10.1145/345124.345171.

149. Petty, R. E., and J. T. Cacioppo. 1984. The effects of involvement on response to argument quantity and quality: central and peripheral routes to persuasion. *Journal of personality and social psychology* 46, no. 1: 69-81.

150. ---. 1986. *Communication and persuasion: central and peripheral routes to attitude change*. New York: Springer-Verlag.

151. Pfaff, G. E. 1985. *User Interface Management Systems*. Secaucus, NJ: Springer-Verlag New York, Inc. .

152. Poole, M. S. 1981. Decision Development in Small Groups I: A Comparison of Two Models. *Communication Monographs* 48, no. 1: 1-24.

153. ---. 1983a. Decision Development in Small Groups II: A Study of Multiple Sequences in Decision Making. *Communication Monographs* 50, no. 3: 206-32.

154. ---. 1983b. Decision development in small groups. III: A multiple sequence model of group decision development. *Communication monographs* 50, no. 4: 321-341.

155. Poole, M. S., and J. A. Doelger. 1986. Development processes in group decision-making. In *Communication and Group Decision-making*, ed. R. Y. Hirokawa and M. S. Poole, 35-62. Sage Publications.

156. Postmes, T., R. Spears, and S. Cihangir. 2001. Quality of decision making and group norms. *Journal of Personality and Social Psychology* 80, no. 6: 918-930.

157. Rich, Charles, and Candace L. Sidner. 1998. COLLAGEN: A Collaboration Manager for Software Interface Agents. *User Modeling and User-Adapted Interaction* 8, no. 3: 315-350. doi:10.1023/A:1008204020038.

158. Rich, E. 1983. Users are individuals:-individualizing user models. *Int. J. Man-Machine Studies* 18: 199-214.

159. Rittel, H. 1972. On the Planning Crisis: Systems Analysis of the "First and Second Generations". *Bedriftsokonomen* 8: 390-396.

160. Rittel, H. W. J., and M. M. Webber. 1973. Dilemmas in a general theory of planning. *Policy Sciences* 4, no. 2 (June 1): 155-169. doi:10.1007/BF01405730.

161. Saaty, T. L. 1990. *Multicriteria Decision Making - The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. RWS Publications.

162. ---. 2001. *The Analytic Network Process*. RWS Publications.

163. Saurí, R., and J. Pustejovsky. 2007. Determining Modality and Factuality for Text Entailment. In , 509-516. September. http://csdl2.computer.org/persagen/DLAbsToc.jsp?resourcePath=/dl/proceedings/icsc/&toc=comp/proceedings/icsc/2007/2997/00/2997toc.xml&DOI=10.1109/ICSC.2007.31.

164. Savage, Leonard J. 1972. *The Foundations of Statistics*. 2nd ed. Dover

Publications, June 1.

165. Schank, R. C. 1972. Conceptual Dependency: A Theory of Natural Language Understanding. *Cognitive Psychology* 3, no. 4: 552-631.

166. Schank, R. C., and R. P. Abelson. 1977. *Scripts, plans, goals, and understanding: an inquiry into human knowledge structures*. Lawrence Erlbaum Associates.

167. Schmidt, K., and C. Simone. 1996. Coordination mechanisms: Towards a conceptual foundation of CSCW systems design. *Computer Supported Cooperative Work (CSCW)* 5, no. 2: 155-200.

168. Searle, J. R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

169. Shackel, Brian. 1991. Usability - context, framework, definition, design and evaluation. In *Human factors for informatics usability*, ed. Brian Shackel and S. J. Richardson, 21-37. Cambridge University Press. http://portal.acm.org/citation.cfm?id=117833.

170. Shafer, G. 1976. *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press .

171. ---. 1988. Savage Revisited. In *Decision Making: Descriptive, Normative, and Prescriptive Interactions*, ed. D. E. Bell, H. Raiffa, and A. Tversky, 193-236. Cambridge University Press.

172. Shafer, G., and R. Logan. 1987. Implementing Dempster's rule for hierarchial evidence. *Artificial Intelligence* 33, no. 3: 271-298.

173. Shafer, G., and A. Tversky. 1988. Languages and Designs for Probability Judgment. In *Decision Making: Descriptive, Normative, and Prescriptive Interactions*, ed. D. E. Bell, H. Raiffa, and A. Tversky, 237-265. Cambridge University Press.

174. Shaw, M. E. 1981. *Group Dynamics: The Psychology of Small Group Behavior.* McGraw-Hill Companies.

175. Shirani, A. I. 2006. Sampling and pooling of decision-relevant information: Comparing the efficiency of face-to-face and GSS supported groups. *Information & Management* 43, no. 4: 521-529.

176. Shneiderman, B. 1983. Direct Manipulation: A Step Beyond Programming Languages. *Computer* 16, no. 8: 57-69.

177. Soller, A. 2001. Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education* 12, no. 1: 40-62.

178. ---. 2004. Computational Modeling and Analysis of Knowledge Sharing in Collaborative Distance Learning. *User Modeling and User-Adapted Interaction* 14, no. 4 (January 1): 351-381. doi:10.1023/B:USER.0000043436.49168.3b.

179. Soller, A., and A. Lesgold. 1999. Analyzing Peer Dialogue from an Active

Learning Perspective. *Proceedings of the AI-ED 99 Workshop: Analysing Educational Dialogue Interaction: Towards Models that Support Learning*: 63Ð71.

180. St. Amant, R., and Paul R. Cohen. 1998. Interaction with a mixed-initiative system for exploratory data analysis. *Knowledge-Based Systems* 10, no. 5 (March): 265-273. doi:10.1016/S0950-7051(97)00038-5.

181. St. Amant, R., M. D. Dinardo, and N. Buckner. 2003. Balancing efficiency and interpretability in an interactive statistical assistant. *Proceedings of the 8th international conference on Intelligent user interfaces*: 181-188.

182. St. Amant, R., and R. Young. 2001. Interface Agents in Model World Environments. *AI Magazine* 22, no. 4: 95-108.

183. Stanton, N. A., M. Young, and B. McCaulder. 1997. Drive-by-wire: The case of driver workload and reclaiming control with adaptive cruise control. *Safety Science* 27, no. 2-3 (December): 149-159. doi:10.1016/S0925-7535(97)00054-4.

184. Stasser, G., and D. Stewart. 1992. Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of Personality and Social Psychology* 63, no. 3: 426-434.

185. Stasser, G., D. D. Stewart, and G. M. Wittenbaum. 1995. Expert Roles and Information Exchange during Discussion: The Importance of Knowing Who Knows What. *Journal of Experimental Social Psychology* 31, no. 3: 244-265.

186. Stasser, G., L. A. Taylor, and C. Hanna. 1989. Information sampling in structured and unstructured discussions of three- and six-person groups. *Journal of personality and social psychology* 57, no. 1: 67-78.

187. Stasser, G., and W. Titus. 1985. Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology* 48, no. 6: 1467-1478.

188. ---. 1987. Effects of information load and percentage of shared information on the dissemination of unshared information during group discussion. *Journal of personality and social psychology* 53, no. 1: 81-93.

189. ---. 2003. Hidden Profiles: A Brief History. *Psychological Inquiry* 14, no. 3&4: 304-313.

190. Steiner, I. D. 1972. *Group processes and productivity*. New York: Academic Press.

191. Stewart, D. D., and G. Stasser. 1998. The sampling of critical, unshared informationin decision-making groups: the role of an informed minority. *European Journal of Social Psychology* 28: 95-113.

192. Straus, S. G. 1996. Getting a Clue: The Effects of Communication Media and Information Distribution on Participation and Performance in Computer-Mediated and Face-to-Face Groups. *Small Group Research* 27, no. 1: 115.

193. Suchman, L. A. 1987. *Plans and situated actions*. Cambridge University Press Cambridge.

194. Suchman, L. A. , and R. H. Trigg. 1992. Understanding practice: video as a medium for reflection and design. In *Design at work: cooperative design of computer systems*, ed. J. Greenbaum and M. Kyng, 65-90. Lawrence Erlbaum Associates, Inc. http://portal.acm.org/citation.cfm?id=125470.125449.

195. Suthers, D. 1999. Representational support for collaborative inquiry. In *HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on System Sciences, 1999.* , Track1:14 pp. doi:10.1109/HICSS.1999.772800.

196. ---. 2003. Representational guidance for collaborative inquiry. In *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments*, ed. J. Andriessen, M. Baker, and D. Suthers, 27-46. Netherlands: Kluwer.

197. Suthers, D., L. Girardeau, and C. Hundhausen. 2003. Deictic roles of external representations in face-to-face and online collaboration. *Designing for change in networked learning environments*: 173-182.

198. Suthers, D., and A. Weiner. 1995. Groupware for developing critical discussion skills. In *Proceedings of Computer Supported Cooperative Learning*, 341-348. Bloomington, Indiana.

199. Suthers, D. D. 2005. Collaborative Knowledge Construction through Shared Representations. *Proceedings of the 38th Hawaii International Conference on the System Sciences (HICSS-38)* Track 1: 5a.

200. Suthers, D. D., R. Vatrapu, R. Medina, S. Joseph, and N. Dwyer. 2008. Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments. *Computers & Education* 50, no. 4: 1103-1127.

201. Terveen, L., M. Stolze, and W. Hill. 1995. From "model world" to "magic world": making graphical objects the medium for intelligent design assistance - a CHI'95 workshop. *ACM SIGCHI Bulletin* 27, no. 4: 31-34.

202. Terveen, L. G. 1995. Overview of human-computer collaboration. *Knowledge-Based Systems* 8, no. 2: 67-81.

203. Toth, E. E., D. D. Suthers, and A. M. Lesgold. 2002. "Mapping to know": The effects of representational guidance and reflective assessment on scientific inquiry. *Science Education* 86, no. 2: 264-286.

204. Toulmin, S. 1958. *The Uses of Argument*. Cambridge University Press.

205. Traum, David Rood. 1995. A computational theory of grounding in natural language conversation. University of Rochester. http://portal.acm.org/citation.cfm?id=220980&dl=.

206. Twardy, C. R. 2004. Argument Maps Improve Critical Thinking. *Teaching Philosophy* 27, no. 2: 95-116.

207. Tyler, S. W., J. L. Schlossberg, and L. K. Cook. 1991. CHORIS: An Intelligent Interface Architecture for Multimodal Interaction. In *AAAI'91 Workshop Notes*, 99-106.

208. Vicente, K. J. 1999. *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*. Lawrence Erlbaum Assoc Inc.

209. Viégas, Fernanda B., Martin Wattenberg, and Kushal Dave. 2004. Studying cooperation and conflict between authors with <i>history flow</i> visualizations. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 575-582. Vienna, Austria: ACM. doi:10.1145/985692.985765. http://portal.acm.org/citation.cfm?id=985692.985765&type=series.

210. Vinokur, A., Y. Trope, and E. Burnstein. 1975. A Decision Making Analysis of Persuasive Argumentation and the Choice Shift Effect. *Journal of Experimental Social Psychology* 11, no. 2: 127-48.

211. Voss, J. F. 2005. Toulmin's Model and the Solving of Ill-Structured Problems. *Argumentation* 19, no. 3 (December 25): 321-329. doi:10.1007/s10503-005-4419-6.

212. Voss, J. F., S. W. Tyler, and L. A. Yengo. 1983. Individual differences in the solving of social science problems. *Individual differences in cognition* 1: 205-232.

213. Winquist, J. R., and J. R. Larson Jr. . 1998. Information pooling: When it impacts group decision making. *Journal of Personality and Social Psychology* 74, no. 2 (February): 371-377.

214. Wittenbaum, G. M., A. Hollingshead, and I. Botero. 2004. From cooperative to motivated information sharing in groups: moving beyond the hidden profile paradigm. *Communication Monographs* 71 (September): 286-310. doi:10.1080/0363452042000299894.

215. Wittenbaum, G. M., A. P. Hubbel, and C. Zuckerman. 1999. Mutual enhancement: Toward an understanding of the collective preference for shared information. *Journal of personality and social psychology* 77, no. 5: 967-978.

216. Wolfman, Steven A., Tessa Lau, Pedro Domingos, and Daniel S. Weld. 2001. Mixed initiative interfaces for learning tasks: SMARTedit talks back. In *Proceedings of the 6th international conference on Intelligent user interfaces*, 167-174. Santa Fe, New Mexico, United States: ACM. doi:10.1145/359784.360332. http://portal.acm.org/citation.cfm?id=359784.360332.

217. Wood, W. 1982. Retrieval of attitude-relevant information from memory: Effects on susceptibility to persuasion and on intrinsic motivation. *Journal of Personality and Social Psychology* 42, no. 5: 798-810.

218. Zadeh, L. A. 1984. Review of Books: A Mathematical Theory of Evidence. *AI Magazine. v5 i3*: 81-83.

219. Zsambok, C. E., and G. A. Klein. 1997. *Naturalistic Decision Making*. Lawrence Erlbaum Associates.

220. Zuber, J. A., H. W. Crott, and J. Werner. 1992. Choice shift and group polarization: An analysis of the status of arguments and social decision schemes. *Journal of Personality and Social Psychology* 62, no. 1: 50-61.