

MAKING SENSE OF STRANGERS' EXPERTISE FROM DIGITAL ARTIFACTS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Nazmus Sadat Shami

January 2009

© 2009 Nazmus Sadat Shami

## MAKING SENSE OF STRANGERS' EXPERTISE FROM DIGITAL ARTIFACTS

Nazmus Sadat Shami, Ph.D.

Cornell University 2009

In organizations, individuals typically rely on their personal networks to obtain expertise when faced with ill-defined problems that require answers that are beyond the scope of their own knowledge. However, individuals cannot always get the needed expertise from their local colleagues. This issue is particularly acute for members in large geographically dispersed organizations since it is difficult to know 'who knows what' among numerous colleagues. The proliferation of social computing technologies such as blogs, online forums, social tags and bookmarks, and social network connection information have expanded the reach and ease at which knowledge workers may become aware of others' expertise. While all these technologies facilitate access to a stranger that can potentially provide needed expertise or advice, there has been little theoretical work on how individuals actually go about this process. I refer to the process of gathering complex, changing and potentially equivocal information, and comprehending it by connecting nuggets of information from many sources to answer vague, non-procedural questions as the process of 'sensemaking'. Through a study of 81 fulltime IBM employees in 21 countries, I look at how existing models and theories of sensemaking and information search may be inadequate to describe the 'people sensemaking' process individuals go through when considering contacting strangers for expertise. Using signaling theory as an interpretive framework, I describe how certain 'signals' in various social software are hard to fake, and are thus more reliable indicators of expertise, approachability, and responsiveness. This research has the

potential to inform models of sensemaking and information search when the search is for people, as opposed to documents.

## BIOGRAPHICAL SKETCH

N. Sadat Shami is a PhD candidate majoring in Human Computer Interaction in the Information Science Program at Cornell University. His research interests are in the areas of social computing, computer mediated communication, and social network analysis. In his research, he draws on theory from the social and behavioral sciences in order to develop a grounded appreciation of how individuals and groups use information technology for information seeking and knowledge sharing. Through such an understanding, he strives to design better systems that support the information seeking and knowledge sharing activities of individuals and groups. Sadat is an active member of the Cornell HCI lab and Cornell SIGCHI. He also interned at the IBM Toronto Software lab and the IBM TJ Watson Research lab in Cambridge, MA. Prior to starting his PhD, Sadat was a research associate at the University of Michigan Ann Arbor, where he developed multi-player online games to study distributed collaboration. His work has been presented at various international conferences such as CHI, CSCW, INTERACT, and HICSS. He is joining the Collaborative User Experience Group at IBM TJ Watson Research Center in Cambridge, MA after graduation.

To my parents  
*Shakila and Shafiqul Islam*

## ACKNOWLEDGMENTS

Many people have encouraged and supported me during my doctoral studies, and on writing this dissertation in particular, and I wish to acknowledge them as a small token of my appreciation.

First, I wish to thank my advisor Geri Gay. During a challenging time of my doctoral studies, Geri kept her faith in me, and remaining calm and confident in my abilities. I would like to acknowledge the critical input of all my committee members, each of whom shared unique insight from their particular field of specialization. Without their constant understanding, informed guidance and enthusiastic encouragement, this dissertation would not have reached its fruition. Their support for my research ideas, even when they remained vague and roughly articulated, was what inspired me forward. Jeff Hancock is a pleasure to work with and amazes me how he can be so intellectually stimulating yet cool to hang out with. Connie Yuan exposed me to much of the essential literature of this dissertation and was instrumental in helping me refine my questions and approach. I took immense pleasure at our dinners at Taste of Thai! Poppy McLeod is a strong mentor, whose perceptive comments as the dreaded ‘reviewer #3’ pushed me to excel. A big thank you to Kate Ehrlich, the best mentor one could ask for! Her unstinting guidance, enthusiasm and encouragement form the bedrock of this dissertation.

I wish to acknowledge IBM Research for allowing me wide-ranging access and flexibility to pursue my dissertation research in their corporate environment. I also want to thank all the IBM employees who gave freely of their time and submitted to being interviewed. In that vein, let me acknowledge Alexandra Jeffery, Amanda Stoller, and Ashley Vernon for their valuable assistance in transcribing said interviews. I acknowledge the feedback and support received from my doctoral

colleagues and many other friends at Cornell. Through offering an intellectually invigorating and personally supportive environment, they have enriched my thinking by sharing their different disciplinary perspectives.

Finally, I acknowledge my parents who gave me the best possible education and start to life, preparing me for this great journey in the first place. Despite the immense distances that separate us, throughout my doctoral years they have offered loving encouragement, support and feedback from afar. Your love and support provided me with the impetus to get done and move on to a new phase of life.



## TABLE OF CONTENTS

<b>BIOGRAPHICAL SKETCH</b>	<b>iii</b>
<b>DEDICATION</b>	<b>iv</b>
<b>ACKNOWLEDGMENTS</b>	<b>v</b>
<b>TABLE OF CONTENTS</b>	<b>vii</b>
<b>LIST OF FIGURES</b>	<b>x</b>
<b>LIST OF TABLES</b>	<b>xi</b>
<b>CHAPTER 1</b>	<b>12</b>
<b>INTRODUCTION</b>	<b>12</b>
<i>Background</i>	12
<i>Prior research on expertise and how people search for it</i>	18
Identifying expertise	26
Selecting Expertise	28
Escalation processes	28
<i>Rethinking the problem of expertise search</i>	29
<i>Applying signaling theory to expertise search</i>	31
<i>Research questions and outline of dissertation</i>	35
<b>CHAPTER 2</b>	<b>42</b>
<b>THEORETICAL FRAMING</b>	<b>42</b>
<i>Looking for experts using technology – an exercise in sensemaking</i>	42
<i>Theories of Information Search</i>	53
<i>Expertise searching and social networks</i>	58
Small world	60
Automatization of network searching	61
Automatization of expertise searching in social networks	62
Important network characteristics that affect network searching	63
<i>Signaling theory</i>	67
The costly to fake principle	70
The full-disclosure principle	74
<i>Signaling theory and its application to human signals</i>	83
<i>Signaling through digital artifacts</i>	89
<i>A preliminary conceptual model of people sensemaking in expertise seeking behavior</i>	92
<b>CHAPTER 3</b>	<b>97</b>
<b>CONTEXT OF STUDY</b>	<b>97</b>
<i>A look at expertise location/recommendation systems</i>	97

<i>Study setting</i>	105
<i>System used</i>	107
<b>CHAPTER 4</b>	<b>111</b>
<b>EMPIRICAL STUDY OF EXPERTISE SEARCH</b>	<b>111</b>
<i>Making sense of initial expertise search results pages</i>	111
<i>Hypotheses</i>	111
<i>Phase 1, Part A: User study with a single keyword</i>	113
Participants	113
Procedure	113
Measures	114
Results	115
<i>Phase 1, Part B: User study with multiple keywords</i>	117
Participants	117
Procedure	117
Measures	118
Results	119
Discussion	120
<i>Making sense of profile pages of experts</i>	121
<i>Signals influencing whom a person decides to contact for expertise</i>	123
Social software as a signal of approachability	123
Social closeness	126
Quality of expertise	127
<i>Phase 2, Part A: User study of profile pages of experts on a single topic</i>	128
Participants	128
Procedure	129
Why AJAX?	131
Measures	132
Quantitative Measures	133
Qualitative Coding	134
Results	136
Participation in social software	137
Social closeness	139
Quality of expertise	140
A note on the role of geography	143
<i>Phase 2, Part B: The role of individual pieces of information within a profile</i>	146
Procedure	148
Qualitative analysis	150
Results	151
Mailing list membership information	152
Social tagging and bookmarking information	152
Social network connection paths	153
Corporate directory information	154
Blog posts	155
Forum posts	155
Self described expertise in corporate directory	156
<i>The role of participation in social software as a proxy of approachability</i>	156

<b>CHAPTER 5</b>	<b>159</b>
<b>CONCLUSION</b>	<b>159</b>
<i>Bringing the pieces together</i>	159
<i>Overview of findings</i>	160
<i>Limitations</i>	165
<i>Contributions to theory and practice</i>	166
Information search theory	166
Signaling theory	169
Implications for practice	172

## LIST OF FIGURES

Figure 1. Master-detail layout in two contexts.	37
Figure 2. Sensemaking process of a Master-Detail page layout.	38
Figure 3. Dervin's sensemaking triangle: Situation-Gap-Help	45
Figure 4. Representation development in Russell et al.'s sensemaking model	48
Figure 5. A preliminary conceptual model of people sensemaking in expertise seeking behavior.	93
Figure 6. Overlap of AJAX experts across different systems	106
Figure 7. The first four of the top ten experts for AJAX.	108
Figure 8. Screenshot of a 'profile' page.	109
Figure 9. Number of times an expert was considered in Phase 1, Part A	116
Figure 10. Number of times an expert was considered in Phase 1, Part B	120
Figure 11. Steps of Phase 2, Part A	131
Figure 12. Social software participation of top ten AJAX experts	132
Figure 13. Reasons behind contacting an expert	135
Figure 14. Reasons behind not contacting an expert	135
Figure 15. Mean social closeness of AJAX experts contacted and not contacted	139
Figure 16. Number of times an expert was considered and contacted in Phase 2, Part A	143
Figure 17. Steps of Phase 2, Part B	149
Figure 18. A preliminary model of 'people sensemaking' in expertise seeking behavior - revisited	169
Figure 19. Screenshot of Facebook interface showing 'mutual contacts'.	173

## LIST OF TABLES

Table 1. A list of tools cited by client facing and non client facing employees as an alternative to SmallBlue.	25
Table 2. Results of GEE for Phase 1, Part A.	116
Table 3. Results of GEE for Phase 1, Part B.	119
Table 4. Results of multi-level regression model for Phase 2, Part A	137
Table 5. Summary of hypotheses and related results.	141
Table 6. Results of multi-level regression model for Phase 2, Part A, when geography is included	146
Table 7. Results of multi-level regression model for Phase 2, Part B	151

## CHAPTER 1

### INTRODUCTION

#### **Background**

Digital artifacts at the same time act as both a rich and an impoverished medium for understanding the expertise of a stranger. On one hand, they can often provide a wealth of information about people that leave behind digital traces through their interaction online. Such information cannot usually be obtained in a face-to-face (FTF) encounter. On the other hand, the credibility and trustworthiness of such information can be difficult to ascertain. As we interact in the digital world, sometimes we intentionally or inadvertently leave behind our digital footprints. We leave behind our footprints when we create a personal homepage, author a blog, or complete our profile on a social networking site. Sometimes information about us is left online by others, as in the case of an online directory or the website of the organization we belong to. These digital traces can be mined by search engines and other purpose built software (Mika, 2005). They may also be aggregated together by data aggregators to automatically create a profile of a person (Ehrlich, Lin, & Griffiths-Fisher, 2007; Lin, Ehrlich, Griffiths-Fisher, & Desforges, 2008; Mika, Elfring, & Groenewegen, 2006).

Digital traces can be utilized to augment and assist the expertise location process in large distributed organizations. In organizations, individuals typically rely on their personal networks to obtain expertise when faced with ill-defined problems that require answers that are beyond the scope of their own knowledge (e.g. Borgatti & Cross, 2003; Cross & Sproull, 2004; Hertzum & Pejtersen, 2000). However, individuals cannot always get the needed expertise from their local colleagues. This issue is particularly acute for members in large geographically dispersed organizations since it is difficult to know ‘who knows what’ among a large number of colleagues

(Constant, Sproull, & Kiesler, 1996). When faced with complex problems that require assistance, individuals have relied on email to ask a group of contacts or a mailing list for expertise (Constant et al., 1996; Weisz, Erickson, & Kellogg, 2006). The expectation is that someone in the distribution list will be able to provide the needed expertise, or forward the email to someone that can. The increasing diffusion and adoption of Web 2.0 technologies have expanded the reach and ease at which the knowledge of others could be utilized. Web 2.0 technologies such as blogs and social tagging and bookmarking emphasize user generated content, interactivity, collaboration and community. These technologies have established a new paradigm of computing and technology development known as 'social computing'. Social computing goes beyond personal computing to facilitate social interactions and collaboration. Participation in such social computing technologies leaves behind digital traces of a person that can be exploited to get an understanding of a person's expertise.

Participating in social computing technologies afford individuals the ability to perform selective self-presentation and impression management (Goffman, 1959). Individuals can portray themselves through personal homepages and social networking profiles as they would like to be perceived. While research on online profiles is clearly emerging, recent findings show that individuals quickly form impressions of personality traits of others from online profiles (Stecher & Counts, 2008). People appear to be able to form accurate impressions of other users' personalities using their profiles. Perceivers' personality trait ratings of Facebook profiles were strongly correlated with users' own self ratings and friends' ratings (Gosling, Gaddis, & Vazire, 2007). Furthermore, users felt that their Facebook profiles could represent them fairly well (Lampe, Ellison, & Steinfield, 2006). However, recent research also shows that there is deception involved in online profiles, raising issues of the

credibility of information found online. Within an online dating context, Hancock, Toma and Ellison demonstrated that deception does occur, albeit in small amounts (Hancock, Toma, & Ellison, 2007).

We are however increasingly noticing information systems that mine content about us which we may not have any control over. People search engines such as Spock<sup>1</sup> and Pipl<sup>2</sup> aggregate both self-authored and other-authored content and present it to anyone using their systems. The content presented through these systems could be content we may not want presented. Essentially, self-authored digital artifacts might differ in content from digital footprints available online that were not created with the intention of self presentation or created by others and are beyond the control of an individual. Gosling et al. call aspects of self presentation in the physical world that one has control over as ‘identity claims’ and ones that occur inadvertently as ‘behavioral residue’ (Gosling, Ko, Mannarelli, & Morris, 2002). Vazire & Gosling extend this to the digital world by demonstrating their existence in digital artifacts (Vazire & Gosling, 2004). This corresponds to Erving Goffman’s distinction between ‘expressions given’ and ‘expressions given off’ (Goffman, 1959). The former are the deliberately transmitted messages intending to show how one wants to be perceived, while the latter are much more unintentional – communicated through nuance and action. With the proliferation of various social computing and search technologies and the ease of sharing information through them, a wide range of information can be available about a person that can be used to draw inferences about him. For example, the impressions formed from looking at self-authored content such as one’s personal homepage may be different from other-authored content such as a blog post about that person.

---

<sup>1</sup> <http://www.spock.com>

<sup>2</sup> <http://pipl.com>



Realizing the affordances of social computing technologies, organizations have aimed to introduce them into the workplace. Large distributed organizations nowadays make available to their employees intranets (Hollingshead, Fulk, & Monge, 2002), blogs (Huh et al., 2007), enhanced corporate directories with people tagging (Farrell, Lau, Nusser, Wilcox, & Muller, 2007), online forums e.g. (Dave, Wattenberg, & Muller, 2004), wikis (Majchrzak, Wagner, & Yates, 2006) and social tagging and bookmarking software (Millen, Feinberg, & Kerr, 2006). Some of these technologies, such as intranets and online forums, have been around for a while. Others, such as blogs, wikis, enhanced corporate directories, and social tagging and bookmarking are more recent developments. Most of these social computing technologies have search capability built into them. These searches return a list of people, and in large distributed organizations, many of whom are unknown to the individual performing the search. Nonetheless, access is provided to a wider range of individuals, making it possible to ask strangers for advice regarding a problem or issue an individual is facing.

There is also software purpose built that allows one to search for or be recommended to experts. Commonly known as ‘expertise locator’ or ‘expertise recommender’ systems, these technologies augment and assist the knowledge discovery process in organizations (See Terveen & McDonald, 2005 for a review). These systems can be thought of as falling into two categories: a) implicit recommender systems, and b) social network based recommender systems. Implicit recommender systems allow individuals to first look for knowledge in documents, and provide pointers to individuals if contact is needed. Answer Garden (Ackerman, 1994), the Designer Assistant (Terveen, Selfridge, & Long, 1995) and PHOAKS (Hill & Terveen, 1996) are examples of systems such as these. They all present relevant information a user searched for, and an email address of the person responsible for the

information in case further contact is needed. On the other hand, social network based expertise recommender systems utilize both expertise information and social connections. Examples of this category of expertise recommender systems are Referral Web (Kautz, Selman, & Shah, 1997a), Expertise Recommender (McDonald & Ackerman, 2000), and SmallBlue<sup>3</sup> (Ehrlich et al., 2007; Lin et al., 2008). ReferralWeb analyzes public web documents to identify names associated with topics, uses co-authorship data to infer social relationships, and presents a referral chain showing the path from the seeker to the expert. Expertise Recommender mines software source control systems and technical support databases to associate specific individuals to specific software modules. It then provides an instant messaging program to users logged into the system to contact individuals with knowledge of the modules. SmallBlue mines outgoing email and instant messaging transcripts and runs a Google PageRank-like algorithm to associate names with topics, as well as to infer social connections.

The increased popularity of various social computing technologies as well as growth of expertise locator systems provides unprecedented levels of awareness and knowledge of others we can interact with. As we interact more often with people who we don't know and have never met in person, we come to rely increasingly on digital artifacts as proxies for directly observable information. We use these digital artifacts to draw rapid inferences about personal characteristics and expected or anticipated behavior that may guide our future interaction (Riegelsberger, Counts, Farnham, & Philips, 2006; Stecher & Counts, 2008). Research has demonstrated that individuals form more exaggerated perceptions of others in the online world (Hancock & Dunham, 2001). In the absence of personal knowledge about a person, it is largely

---

<sup>3</sup> Later renamed to Atlas™

perceptions and inferences that dictate whom a person contacts for specific expertise (Hinds, Carley, Krackhardt, & Wholey, 2000).

In the milieu of proliferating digital information about individuals, it is not the lack of information, but which information one should pay attention to that becomes the challenge. The vast volume of online information, the varying degrees of validity of such information, and its often non-relevance to the question at hand may overwhelm individuals. This is particularly critical when we seek expertise from others based on perceptions of digital information. Technology mediated expertise search is largely about searching amongst strangers since most people will turn first to the people they know to get needed information (Borgatti & Cross, 2003; Hertzum & Pejtersen, 2000) and only later use tools to seek out experts. This makes expertise search a good task for exploring issues of perceptions of information about strangers since there is a clear purpose to the interpretation.

Seeking to contact others for expertise using technology involves a set of interconnected cognitive activities, including generating a query, searching for relevant information, evaluating and making sense of information found, and coherently integrating different pieces information into a coherent whole to arrive at a decision. It may involve sifting through massive volumes of information under deadline pressure to make complex search decisions under uncertain conditions. The process of gathering complex, changing and potentially equivocal information, and comprehending it by *connecting nuggets of information* from *many sources* to answer vague, *non-procedural* questions is known as sensemaking (Dervin, 1992; Gotz, 2007; Russell, Stefik, Pirolli, & Card, 1993; Weick, 1995). Although there is some confusion regarding what exactly constitutes sensemaking, as suggested by recently published articles with names such as ‘Making sense of sensemaking’ (Furnas & Russell, 2005; G. Klein, Moon, & Hoffman, 2006a, 2006b; Whittaker, 2008), in my

dissertation I will use the above definition of sensemaking. In the sensemaking process, individuals do not rely on a single source of information. Rather, they integrate multiple sources of information and synthesize that information together. There is also no fixed procedure in place for a sensemaking task such as looking for experts using technology. This is in contrast to a procedural task such as purchasing a book online. To purchase a book online, a user adds it to her shopping cart, and follows the typical checkout procedures by entering credit card information, and shipping and billing addresses.

Searching for experts using technology could thus be considered an exercise in sensemaking (Gotz, 2007; Russell et al., 1993). A theory that can inform this sensemaking process is signaling theory. This dissertation will explore how signaling theory informs the process of making sense of strangers' expertise from digital artifacts. But before that, a discussion on prior research on using technology to search for experts seems relevant.

### **Prior research on expertise and how people search for it**

Expertise is defined differently in different disciplines. In the field of psychology, expertise is defined as a human cognitive *skill* acquired by repeatedly performing a task (Anderson, 2000). People who have a kind of expertise in a particular topic are called experts. Many early expert databases systems were designed according to this definition. The experts who input into the database are publicly recognized people who are the best (or close to the best) in a certain domain. However, according to this definition, few people can claim themselves as experts in reality, although most will agree that they have expertise in some areas. In many knowledge seeking tasks, finding a person with sufficient expertise instead of an optimal expert is a more practical solution. Depending on the task at hand, finding an optimal expert may be ideal. But it is also much more difficult since the optimal

expert may not respond to an expertise query. This is similar to the idea proposed by Simon (1972) – people seldom make fully informed decisions but rather ‘satisfice’.

This dissertation deals with the processes through which people make sense of available digital information in searching for experts. The way people go about looking for experts may emphasize making use of locally available expertise instead of finding the optimal one. Thus, in this dissertation, I adopted a more practical view of expertise proposed by Ackerman and Halverson (2004), in which “expertise connotes relative levels of knowledge in people”. According to this definition, expertise is a range and an individual can have different levels of expertise on different topics. Such expertise is arranged and valued by the social and organizational settings where the individuals are evaluated.

Expertise is available in a variety of sources. Expertise can be obtained from humans, as well as non-human sources such as books and webpages. Yuan, Fulk and Monge (2007) draw the distinction between connective and communal knowledge sources. Connective knowledge sources represent human experts while communal knowledge sources represent digital knowledge repositories. When organization members cannot easily locate connective knowledge sources, they may turn to communal knowledge sources for expertise, as long as they perceive such sources to contain the expertise sought (Hollingshead et al., 2002). Yuan and colleagues also found that individuals’ retrieval from digital knowledge repositories was positively related to their contribution to the repository (Yuan et al., 2005). Through two case studies, Hertzum & Pejtersen analyzed the factors that influence engineers’ choice of information sources (Hertzum & Pejtersen, 2000). They found that the choice between choosing documents and choosing people for expertise was a function of task characteristics. Fidel & Green performed a somewhat similar study in which they looked at the circumstances in which engineers selected human sources and

circumstances in which they selected documents (Fidel & Green, 2004). Woudstra & van den Hooff performed an experimental study where they had participants look at online profiles of fellow employees on a corporate intranet. The profiles contained contact details, information about the person's knowledge area(s), job position and educational background (Woudstra & van den Hooff, 2008).

Expertise sharing is viewed as the next step of knowledge management for organizations by many scholars (e.g. Ackerman & Halverson, 2003). First generation knowledge management focused on a repository approach of using information technology to manage organizational knowledge (Ackerman, Wulf, & Pipek, 2002). Its key idea was to externalize knowledge from individuals and place it into shared repositories, such as an information database or knowledge base, as documents for later retrieval and use. Its theoretical foundation was a "knowledge creation model" proposed by Nonaka and Takeuchi (1995). In this model, knowledge creation is a spiraling process of interactions between explicit and tacit knowledge, which includes processes of socialization, externalization, combination, and internalization of knowledge. Based on this model, knowledge management systems tend to emphasize gathering, storing, providing, and filtering available explicit knowledge. Such repository view of knowledge management has its advantages. By using standard technology and controlled input, the information put into the repository is easy to search, access, and transfer. By externalizing individuals' knowledge, it also makes organizations less vulnerable to employee turnover (Argote, 1999). However, this approach is limited and is difficult to apply in some situations. For instance, Lave and Wenger (1991) suggested that expertise is usually embedded in some particular situations and environments and is hard to extract. Hinds and Pfeffer (2003) found that it is difficult for people to use the de-contextualized information that is stored in the knowledge base as well as transfer the same knowledge into other contexts.

Expertise sharing aims to help people share their expertise, to provide information seekers access to knowledge held by people directly, which complements the limitations of accessing information from documents. For instance, by enabling two-way interactions between seekers and experts, it is easier for people to build common ground, understand the asker's context and needs, and transfer tacit knowledge. By not requiring experts to totally externalize their knowledge but instead help others in a case by case basis, it may also make them less concerned with losing their power (Hinds & Pfeffer, 2003).

Although there are many benefits from seeking information from people directly, in reality, people are not always the first choice for information seekers. Research has found that there are various barriers for people seeking expertise from their colleagues, including social costs and logistical costs (i.e. easy access to the source). With the wide adoption of advanced communication technology, perhaps the logistical barriers to reach other people will be less difficult in the future. Thus, here I focus the discussion on the related social costs.

In his study conducted in an industrial lab, Allen (1977) noted that engineers approached their colleagues less frequently as their first resource for information compared with documented literature, although they agreed that their colleagues could provide high-quality information. Allen found that the major factor affecting peoples' searching source selection is accessibility. A person considered to be an expert might not respond to an expertise query. Allen indicated that, compared to searching for and reading literature, asking help from colleagues has psychological costs as well, which include the potential lack of reciprocity between giving and obtaining information, as well as the status implications of admitting ignorance. This social psychological cost seems to outweigh the benefits of consulting people directly. For instance, Allen found that even when they needed to consult their colleagues, engineers tend to go to

the literature first to improve their background in the area so they will not appear ignorant.

Similar findings can be found in later studies in the field of social psychology. Lee (2002) found that in an organization, fewer than one-third of participants who needed help to solve a problem proactively asked other people for help, even though help was available. Lee found that this is because the social cost, including admitting incompetence, inferiority, and dependence, is expensive for a help seeker as it hurts self-esteem and public impression. Furthermore, DePaulo and Fisher (1980) found that a person deciding whether to ask for help not only takes into account his own costs, but also the “anticipated cost-reward contingencies” of the helper. An excellent review of various factors that affect people help-seeking behavior can be found in Gall (1985).

It is noteworthy that the social psychological costs for asking for informational help are fluid and vary in different circumstances. Allen (1977) found that developing social relationships is an effective strategy to decrease the concerns of social psychological cost. When information seekers have good social relationships with available helpers, they tend to worry less about the social cost and can communicate more effectively. The benefit of using social relationships to seek help can also be found in the social network literature (Haythornthwaite, 2002; Shapiro, 1980). Furthermore, Lee (2002) found that the social cost of help seeking is lower for peripheral tasks than central tasks. This implies that when the expertise sought is not related to something that could be considered a reflection of a person’s competence (such as a graphics designer asking about how to program a microcontroller), the social cost may not be as important to them.

These social costs raise an interesting research question regarding how people choose whom to contact for specific expertise. Yimam-Seid and Kobsa (2003)



divided the needs of people searching for expertise into two categories: a) looking for a person as a source of information and b) looking for someone who can perform a given organizational or social function, such as giving a speech. Yimam-Seid and Kobsa (2003) suggested that there are different reasons for people choosing a person over other sources. The major ones include:

- Accessing undocumented or nonpublic information. Not all information is accessible because of different cognitive, economic, social, or political reasons (Kautz et al., 1997a).
- Solving problems that are situated. For instance, Orr (1996) showed how informal interpersonal interactions in the form of narratives lead individuals to new understandings of work related problems.
- Leveraging others' expertise to minimize the time and effort in information seeking. For many information seeking tasks, it may take a lot of work for novices but only a little work for experts, especially when people search for information in areas in which they are not familiar (Bhavnani, 2005). Experts can help users quickly formulate their information needs into query terms and point them to the valuable information sources available without spending much time (Taylor, 1962).

Ehrlich and Shami (2008) extend the work of Yimam-Seid and Kobsa (2003) through an empirical study of how individuals go about searching for people for expertise. They found that when using information retrieval systems to search for experts, people perform four types of queries. These are queries for 1) finding answers, 2) finding people, 3) awareness, and 4) providing information. Finding answers refers to getting an answer to a specific question where the answer is more important than who answers it and does not require 2-way discussion. For example, the search term 'camtasia' was used to find out to record using Camtasia software.

This type of information is usually located in documentation and does not require follow-up with a person. On the other hand, finding people is when the need is to find a person with specific skills. For example, the search term ‘ruby programming’ was entered to find someone that had experience with the Ruby programming language. The information seeker had to deploy a ruby application for a client in Japan and did not have experience with Ruby.

While ‘finding answers’ and ‘finding people’ have been reported in Yimam-Seid and Kobsa (2003), the categories of ‘awareness’ and ‘providing information’ were unreported. ‘Awareness’ queries involve developing knowledge of a topic where neither the topic nor the person is specific. The example of a user looking up the term ‘health medical records’ is given where the employee is located in a small country and wants to get an idea of who else is doing work in that area nearby. Finally, ‘providing information’ queries were defined as the seeker having information that might be valuable and wants to find others that could use it. The authors provide an example of an employee that did a search on ‘workforce and mining industry’ so he could share the experience he gathered from visiting with 10 mining companies around their workforce issues in Australia. This showed that employees use search systems not just to seek expertise from others but to provide it to others as well.

In addition to outlining the reasons that people look for experts using technology, Ehrlich and Shami (2008) provide a brief overview of the range of different systems used in this process and how their use varies by job function, namely client-facing and non-client facing. Many of these systems could be considered information retrieval systems since they have search capabilities built into them. Table 1 shows this list of systems as well as how often they are used by client-facing and non-client facing employees in the organization they studied. The table

illustrates the wide breadth of technologies used within an organization to look for expertise.

Table 1. A list of tools cited by client facing and non client facing employees as an alternative to SmallBlue.

<b>TOOL</b>	<b>Not Client-facing</b>	<b>Client-facing</b>
<b>Directory.</b> An enterprise directory populated with information from HR records augmented with additional information provided by the individual	33%	44%
<b>Personal network.</b>	23%	29%
<b>Profile.</b> Dynamic profile ( e.g. Farrell, Lau, Nusser et al., 2007)	15%	5%
<b>Broadcast.</b> Tools, including email, for sending broadcast requests. Purpose built tools e.g. (Weisz, Erickson, & Kellogg 2006) were not widely available at the time of the study	15%	6%
<b>Intranet.</b>	8%	10%
<b>Social software.</b> Searching user generated data such as blogs, wikis, and enterprise social bookmarking software	6%	3%
<b>Documents.</b> An online repository	0%	3%

Penuel and Cohen (2003) pointed out that the need of expertise is also related to individual experience. They found that there are two different types of knowledge learning needs in organizations: the learning of newcomers or novices on the job, and the learning of experts. They have different backgrounds and need different supporting strategies. For a newcomer, the most important thing is to find out where expertise is distributed and how they can access it. For experts, they may already know these things, and their needs may be more related to interaction with other experts or people to update and expand their knowledge or solve new problems. In summary, the

literature reviewed above illustrate how variegated and situated peoples' expertise needs are as well as the diverse set of techniques and technologies people use to satisfy their expertise needs.

McDonald and Ackerman (1998) provide one of the earliest studies to systemically investigate how people search for expertise in organizations. They suggested that the process of finding expertise includes three steps: "expertise identification", "expertise selection", and "escalation processes". In following subsections, I use this framework, combined with other related studies, to discuss how people search for expertise in organizations.

### *Identifying expertise*

Expertise identification is about "knowing what information or special skills other individuals have" (McDonald & Ackerman, 1998). It is the first crucial step in the process of expertise searching. Understanding how people identify expertise in real life can help us understand how to augment this process in system design. McDonald and Ackerman (1998) found three ways in which people identify expertise: everyday expertise, historical artifacts, and expertise concierges.

"Everyday expertise" is about knowing who knows what by everyday "experience." Similar findings can be found in the studies of "transactive memory" (Moreland, Argote, & Krishnan, 1996; Wegner, 1987, 1995). The key idea is that people get to know their colleagues' expertise based on their daily interactions. "Everyday expertise" is affected by people's professional experience, organizational tenure, and geographical proximity.

"Historical artifacts" are archival data such as software source code change log history, which can indicate one's previous work and related expertise.

"Expertise concierges" are about using some specific people who know others well to refer information-seekers to the possible helpers. This concept is similar to

“technological gatekeepers” described in Allen (1977) and “contact brokers” described in Paepcke (1996). In organizations, these are people who usually have strong social networks. They maintain “a sophisticated map of the individuals in the organization and what they know” (McDonald & Ackerman, 1998). They play the role that mediates information-seeking requests to those who are most likely to have the information. In their study, McDonald and Ackerman noted that these people are usually managers, who have a “high level of technical competence” and “relatively long tenure with the organization” and “high-status positions.”

Another interesting work on how people get to know one another’s expertise is Fitzpatrick’s case study in a new community. Fitzpatrick (2003) summarized how people get to know others by “finding out in the large” and “finding out in the small”. Information “in the large” is that information “of relatively coarse grain and likely to be easy to find out... People are more likely to self-report or that is more amenable to being recorded in some form or to being publicly available” (p. 92). Such information includes who worked on what and who knew whom. Fitzpatrick (2003) found that people are likely to gain such information through previous experience or from general conversation. Information “in the small” is that “information which is at a much finer level of granularity that people would rarely think to self-report because they would not deem it relevant or important at the time” (p. 93), such as shortcuts to do a specific task. Such information is usually discovered and shared “by accident in the course of casual conversation”, such as “finding out accidentally, finding out by snooping, finding out incidentally, finding out incrementally, and finding out the real story” (p. 94).

In summary, although the task of searching for experts takes place in a relatively short time span, the process of knowing where experts are located is situated in people’s everyday activities, including their experience, social interactions, and

artifacts.

### ***Selecting Expertise***

Expertise seekers usually are faced with choosing from among several possible alternatives that have the needed expertise. However, to augment this process, we need to understand what criteria are important.

As mentioned, similar social costs (i.e. loss of status), expected reciprocity (i.e. can I return the favor later) and social equity (i.e. how well do they know each other socially) are the key factors that affect decisions on whom to ask for expertise (Allen, 1977). Lee (2002) found that people prefer to seek help from peers instead of higher or lower levels of their organization's hierarchy because of such social cost considerations.

McDonald and Ackerman (1998) further explored the expertise selection problem in detail. They identified three general expertise selection mechanisms: organizational criteria, the load on the source, and performance. Their findings include that people tend to go to local experts first, they compare expert candidates' workload (both regular and over time) before going to them, and they consider an expert's ability for problem comprehension and providing a suitable explanation, as well as their attitude.

In summary, we can see the social and psychological complexity of the expertise selection problem. As McDonald and Ackerman (1998) summarized, "expertise selection is achieved through combinations of many, slightly different, behaviors each adding to an individual's judgment about the appropriateness of one or more expertise candidates" (p. 320).

### ***Escalation processes***

Finally, McDonald and Ackerman also indicated that expertise finding often involves escalation processes. Escalation is "the way in which people repair failures

in identification and selection” (McDonald & Ackerman, 1998, p. 322). Expertise identification can fail in three ways: over-identification (the set of candidates provided is too large), under-identification (the set of candidates provided is too small), or misidentification (none of the candidates provided has the required expertise at a sufficient level). Expertise selection can fail when the selected expert is too busy to respond or does not really understand the problem. McDonald and Ackerman pointed out that escalation provides a way to either adjust the set of candidates previously identified or to reselect from among those candidates utilizing information gained in the previous attempts. They suggest that expertise location systems should support such escalation process, such as having some feedback and modification techniques to support users’ previous histories or personal preferences.

### **Rethinking the problem of expertise search**

Given the discussion of prior research on expertise search in the previous section, there is a need to rethink how to approach research in this area. As mentioned previously, there has been a shift from attempting to capture people’s knowledge in digital repositories to identifying people with knowledge. There has been a shift from identifying the best person possible to identifying someone who might be able to provide a reasonable answer or point to someone that can. We may need to rethink the notion of expertise and frame it as a trait rather than a skill that can be quantified. This might be akin to looking for someone with attractive personality traits. These traits tend to be things that are inferred rather explicit.

Compared to seeking information from a library or the web, searching for expertise from people has many unique benefits. However, it also raises many issues socially, such as various expertise needs and the associated costs for seekers. Although the expertise searching task seems to take place in only a short time, from the analysis of people’s search for expertise in organizations, we can see that it is

tightly coupled with an individual's social experience and organizational structure and culture.

Based on the literature reviewed on expertise search in the prior section, the following issues appear to be worthy of research.

1) To consider and support various ways to identify expertise from different types of artifacts: Using historical artifacts is a practical way of identifying expertise. We automatically create electronic records (e.g. source code, emails, etc.) during the course of doing everyday work. Through the development of information retrieval technology, we can easily mine these digital artifacts to find out what people created or accessed, which hints at what people know or are good at.

There has also been increasing use of social computing technologies such as blogs, forums, social tagging and bookmarking, and social networking sites. These can also be mined and mashed up together to create a profile of a person (Mika et al., 2006). Previous systems have not put much attention into the consideration of the social network structure that underlies the expertise searching and accessing process. Newer systems, such as Expertise Recommender (McDonald & Ackerman, 2000) and SmalBlue (Ehrlich et al., 2007; Lin et al., 2008), have started to look at using social networks as a means of searching and accessing people. However, there has not been any systematic analysis on how systems with the consideration of social computing technologies such as blogs, social tags and bookmarks, forums and social network characteristics are used by individuals. Additionally, some systems (e.g. SmallBlue) provide a person's role and their position in the organizational hierarchical structure, which indirectly reflect one's experience. There is also a lack of study on how this "extra" information affects people's usage of the system.

2) To consider various factors that affect people's decisions on expertise selection. Identifying experts is not the end of the expertise searching process.



Simply giving people the best expert available may not work. A more preferred way is providing information seekers candidates who have a satisfying (instead of best) expertise but a low social cost to access.

It is difficult (if not impossible) to implement one “identifying and selecting algorithm” for expertise selection (Zhang & Ackerman, 2005). As McDonald and Ackerman (1998) pointed out, systems should not automatically select an expert for information seekers. Instead, it should provide a list of candidates with related information to support people’s decision making. There should be cues about social connection, availability, position within the organization etc.

Good social relationships can decrease the social cost of expertise searching. The process of asking and answering questions is also a process of using and building social relationships. It is an interesting research endeavor to see how people balance the need to obtain good expertise with the ease of accessing expertise through existing social connections.

The concept of “expertise concierges” is worthy of being operationalized and further studied within expertise locator systems. This is an extremely important method for people to find possible helpers outside of their immediate social environment or daily experience. A key research issue is how to identify these expertise concierges and make them more accessible to people.

### **Applying signaling theory to expertise search**

Signaling theory provides a useful framework in understanding which pieces of information may be more reliable when making inferences regarding a person’s expertise. Reliable signals are pieces of information that are hard to fake. Such information allows users to ‘separate the grain from the chaff’ by distinguishing between different types of information. For example, social network connection

information may be a more reliable signal of expertise because people within a social network connection chain can credential the expertise of an individual.

Signaling theory has its origins in both economics and biology. Spence described an economic theoretical framework for signaling (Spence, 1973). Employers, lacking direct information about prospective employees' productivity, use market signals to improve the chances of hiring productive employees. Spence defined signals as personal attributes, such as education and work experience, that an individual can change. Individuals make choices about investments in education in order to maximize the difference between their educational expenses, or signaling costs, and the wages offered by employers. Education does not necessarily improve an individual's capacities or raise their productivity. Rather, it is a screening device that functions to identify individuals with innate characteristics that make them more productive. A prospective employee's level of education serves as a signal to the employer regarding his or her likely productivity.

In biology, signaling theory has been used to explain seemingly wasteful and detrimental ornaments and behaviors in animals (Zahavi, 1975). The signal itself, carried in behaviors and other phenotypic traits, is costly in terms of time, energy, or risk, making it difficult to fake, and ensuring that the signal transmits reliable information to the signal receivers. Among the frequently cited examples of costly signal use in predator deterrence is stotting in gazelles (Zahavi, 1975). When a gazelle notices a predator, the gazelle stomps its feet and turns away from the predator, showing a black and white rump. Then the gazelle will stot, jumping high into the air on all four legs. Although this behavior reveals the gazelle to the predator, it also serves as a reliable signal that the gazelle is in good physical condition and is likely to outrun the predator if pursued. Because stotting requires great energetic expenditure and wasteful use of valuable "escape time," only gazelles that are in good condition

will stot. For this reason, stotting is a reliable signal to the predator that a long, difficult pursuit will only result in failure and exhaustion. Zahavi and Zahavi have identified several means by which animals signal to competitors, including singing, aerial display, electric pulses, posturing, and the release of chemicals (Zahavi & Zahavi, 1997). Physical attributes can also serve as honest signals of quality to attract potential mates. For instance, bright coloration in males is an honest signal of quality because it is likely to attract the attention of predators. These colors may also attract female attention to size, shape, and movement of the males (Zahavi & Zahavi, 1997). Those males who are able to survive with these bright colors may be higher in quality and more desirable. Common examples include the massive tail feathers of male peacocks. These characteristics make the male more vulnerable to recognition and attack by predators, and require strong physical constitution and adequate nutrition. For these reasons, ornaments can be honest, reliable signals of quality in a mate.

Judith Donath talks about three types of signals in digital artifacts: 1) handicap signals, 2) index signals, and, 3) conventional signals (Donath, In Press). *Handicap signals* are costly to produce and are considered reliable because the quality they signal is ‘wasted’ in the production of the signal, and the signal tends to be more expensive to produce for an individual with less of the quality. An example of a handicap signal is active participation in online forums. An employee with over 10,000 forum posts proves that she has enough time to be active in the forum, while still maintaining her job responsibilities. She ‘wastes’ time to prove she has a surplus. “The Handicap Principle is a very simple idea: waste can make sense, because by wasting one proves conclusively that one has enough assets to waste and more” (Zahavi & Zahavi, 1997).

*Index signals* are directly related to the trait being advertised. These are reliable since they require that the sender possesses the relevant trait. For example,

being a level 60 avatar, with accompanying powerful sets of armor and weapons in the popular multi-player online game World of Warcraft is an index signal. Having the quality of being a good gamer is a pre-requisite to produce this signal. This connection between signal and quality makes an index signal reliable. Handicap and index signals are known together as *assessment signals*. Assessment signals relate to the quality it represents and thus one can assess the quality simply by observing the signal (Donath, In Press).

On the other hand, *conventional signals* are not correlated with a trait. The signaler need not possess the trait to send the signal. Because of this, conventional signals are less reliable and open to deception. The online world is rife with conventional signals. For example, it may be desirable to have an attractive picture of oneself on a social networking site such as MySpace. In the absence of social connections that can vouch for the veracity of such a picture, an individual may choose to put up a deceptive picture. If the use of such deceptive pictures becomes prevalent, the signal will lose its meaning as an indicator of attractiveness. Conventional signals are thus unstable because excessive deception can cause a once meaningful signal to turn into noise (Donath, 1999).

Signaling theory proposes that there are costs and benefits to both the sender of the signal and the receiver. For example, research has found that humans sometimes form automatic impressions on the basis of prior experiences (Greenwald & Banaji, 1995). However, when looking for expertise, one may want to engage in more detailed processing. Signaling theory provides an explanation regarding situations in which individuals may engage in automatic processing versus detailed processing. There is a concept of 'receiver costs' in signaling theory. If a reliable signal is very costly to assess, receivers might choose one that is less reliable but easier to obtain (Guilford & Dawkins, 1991). When the cost of making a poor decision is great,

individuals will spend more time evaluating reliable signals and less time making automatic inferences. When the task at hand does not involve a high cost if a poor decision is made, individuals may engage in satisficing behaviors through automatic processing.

Given the proliferation of conventional signals online, it is not surprising that the majority of research on signaling in digital artifacts has been related to that type of signal. Donath looked at signaling in social networking sites such as Friendster and MySpace (Donath, 2007), where one might potentially artificially inflate his friends to appear popular or because of the social pressure to accept friend requests. Lampe, Ellison and Steinfield looked at another social networking site where users can selectively self-present themselves (Lampe et al., 2006). Investigating student behavior in the popular social networking site Facebook, they found that the completion of particular profile fields was a strong predictor of how many friends a student had. However, in the online world, assessment signals could be juxtaposed with conventional signals, albeit to a lesser degree. Inferred social connection information, as opposed to self reported social connection information which could potentially be deceptive, may act as an assessment signal of one's sociability. In a similar way, expertise rank in an expertise locator system that is determined through an algorithmic process may act as an assessment signal of expertise. A contribution of this dissertation is to look at *both* assessment signals and conventional signals and how they are perceived.

### **Research questions and outline of dissertation**

This dissertation will attempt to elucidate how people form impressions of a person's expertise from the digital artifacts available about them during an expertise searching activity. In the majority of cases, these people will be strangers since one

would not be using technology to search for a person she already knows has the needed expertise. She would just contact that expert directly.

This expertise searching activity discussed in this dissertation will be mediated through an information retrieval system. Searching for experts using technology could be thought of to consist of two distinct phases. The first part is the typical enter query/review results approach, which has been studied extensively (e.g. Granka, Joachims, & Gay, 2004; Joachims, Granka, Pan, & Gay, 2005; Joachims et al., 2007; Pan et al., 2007). The second phase is where the bulk of this dissertation is concerned with. This phase involves disambiguating the context of the identified experts to gauge factors beyond expertise such as availability, responsiveness and credibility. While performing the various types of expertise queries outlined in Ehrlich and Shami (2008), individuals information seeking behavior is largely shaped by the structure of the user interface, that is, the information environment. In this dissertation, I will thus be looking at information retrieval systems that follow a Master-Detail page layout. A Master-Detail page layout is one of the most common user interface displays for presenting search results (Muck & West, 2004). Within this layout, a ‘Master’ page contains a list of search results, with each search result containing metadata and/or summary information about that result. Once a user clicks on a search result, it takes them to the actual web page. Examples of Master-Detail page architectures in two contexts are displayed in Figure 1. Master-detail page architectures are prevalent in a wide range of search interfaces ranging from product searches on sites such as Amazon<sup>4</sup> to name searches on social networking sites such as Facebook<sup>5</sup> and MySpace<sup>6</sup>. While emerging technology such as AJAX allows one to view previews of

---

<sup>4</sup> <http://www.amazon.com>

<sup>5</sup> <http://www.facebook.com>

<sup>6</sup> <http://www.myspace.com>

Detail pages without leaving the Master page (e.g. the Netflix<sup>7</sup> interface), the majority of web based search applications still follow the Master-Detail page layout.



Figure 1. Master-detail layout in two contexts. (A) is a master page and (a1) and (a2) are detail pages when looking up a person in the Google search engine. (B) is a master page and (b1) a detail page when searching for people with matching interests in the popular social networking site MySpace.

<sup>7</sup> <http://www.netflix.com>

We can think of the Master-Detail page layout as a cyclical process of making sense of information. A user types in a query term and sees a list of any  $n$  number of search results displayed on the Master page. The user tries to make sense of the information on the Master page. If she is not satisfied with the results on the Master page, she may reformulate the query. Otherwise she may choose to explore any  $n$  number of detail pages, depending on how satisfied she is with each of them. This iterative process is outlined in Figure 2.

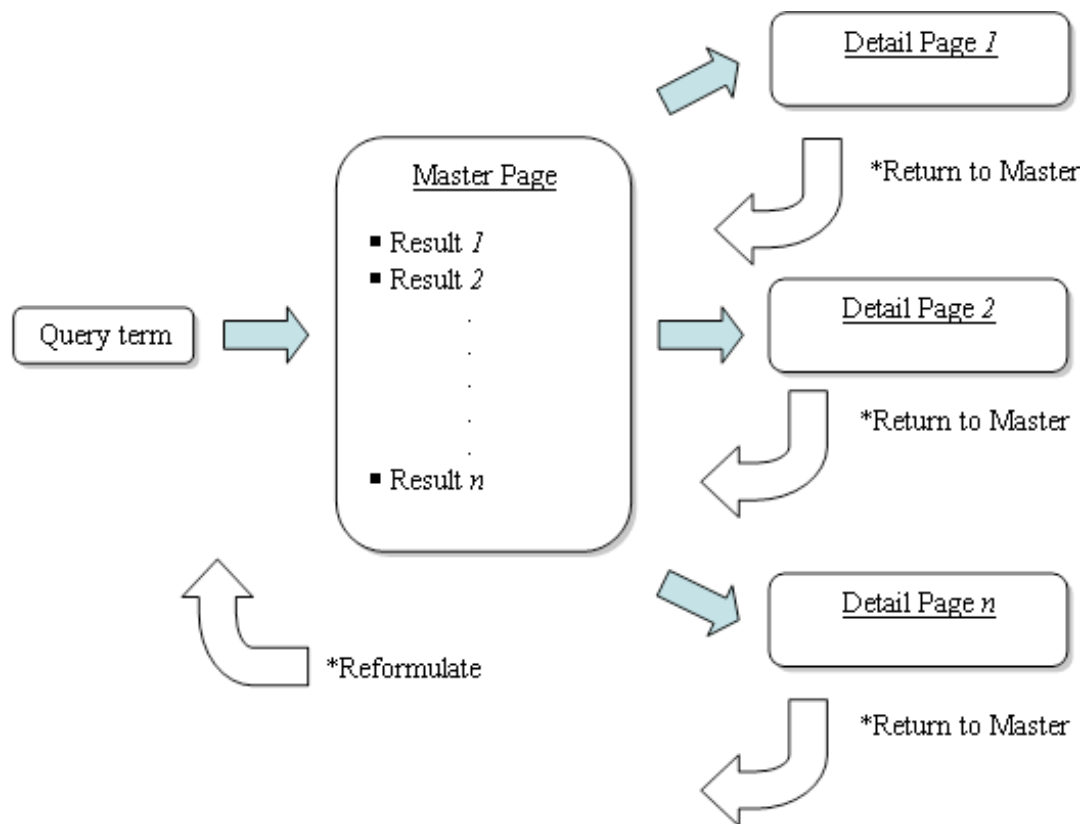


Figure 2. Sensemaking process of a Master-Detail page layout. (\*) indicates situations in which the user is not satisfied with the results.

Within the context of this search interface, my dissertation examines the following research questions:



1. How do people who have passing knowledge on a topic make sense of the information on an initial search result page (Master page) when seeking to contact an expert on that topic? What factors influence their decision? This will be discussed in phase 1 of the study described in chapter 4, where participants search for an expert in a topic that is assigned by the experimenter.
2. How do people make sense of the information on a search result page (Master page) when they have considerable expertise in the skill they are performing the search for? Which factors are important in deciding to click on a particular search result for further exploration? This will also be discussed in phase 1 of the study described in chapter 4, where participants search for an expert on a topic of their choosing.
3. How do people weigh the various pieces of information on an expert's profile page (Detail page) that have been aggregated together from various data sources? How do people form impressions of factors such as availability and accessibility? This will be discussed in detail in phase 2 of the study in chapter 4.

In order to address these questions, I will combine quantitative model building with qualitative data analysis. The remainder of this dissertation is thus organized as follows. Chapter 2 will provide theoretical grounding for the dissertation. It will review existing theories on information search, sensemaking, and relevant research from the social network analysis literature. It will then propose a preliminary model of 'people sensemaking' based on signaling theory.

Chapter 3 will provide a review of existing expertise locator/recommender systems and provide the rationale behind using the particular expertise locator system used in this dissertation. The preliminary conceptual model of 'people sensemaking'

developed in chapter 2 will be applied to elucidate expertise seeking behavior using this system. The system chosen was the SmallBlue expertise locator system (Ehrlich et al., 2007; Lin et al., 2008). This system was chosen because it is an ideal test bed for the research questions under investigation.

Chapter 4 will detail the empirical study carried out for this dissertation. Phase 1 of the study will describe how individuals make sense of a search result page (Master page). The degree of prior knowledge in the expertise keyword that is searched might influence how individuals make sense of a search result page of experts. Using the search results from the SmallBlue expertise locator system (Ehrlich et al., 2007; Lin et al., 2008) as the basis of analysis, phase 1 will focus on how individuals make sense of the summary information in the initial search result page when they a) lack knowledge, and b) are knowledgeable about the expertise they are seeking. A typical search result page displays search results in a rank ordered manner with summary information about each search result. Prior studies in document search have shown that rank order matters in which search result is selected for further exploration. Research has found that higher ranked results are selected significantly more than lower ranked results (Granka et al., 2004; Joachims et al., 2007; Pan et al., 2007). Does this pattern also hold for search results of people search? In addition to rank order, the search results of the SmallBlue system displays signals regarding the social relationship between the seeker and the target. Phase 1 will explore the role of these social connection signals in how they influence the decision of which search result is selected for further exploration. The two scenarios of phase 1 will look at the effect of rank order and social connection information, but will vary the prior knowledge of the seeker in the expertise being sought. In the first scenario, participants will have passing knowledge of the search term, while in the second scenario they will have considerable knowledge in the expertise being sought.

Chapter 4 will also detail phase 2 of the study. This phase will be dedicated to applying the preliminary conceptual model of ‘people sensemaking’ to how individuals make sense of the profile page (Detail page) of an expert. Through the profile page of an expert in SmallBlue, I will investigate which signals are more reliable predictors of whom a person eventually decides to contact for specific expertise. Employing a qualitative lens, I will also elaborate on the explanatory power of the conceptual model of ‘people sensemaking’ by illustrating how certain signals may be more reliable than others.

In chapter 5, I will conclude with a summary of theoretical and design implications from my study, limitations of the study, and future research directions.

## CHAPTER 2

### THEORETICAL FRAMING

#### **Looking for experts using technology – an exercise in sensemaking**

As previously discussed, seeking to contact others for expertise using technology involves a set of interconnected cognitive activities, including generating a query, searching for relevant information, evaluating and making sense of information found, and coherently integrating different pieces of information into a coherent whole to arrive at a decision. We can define this process of gathering complex, changing and potentially equivocal information, and comprehending it by connecting nuggets of information from many sources to answer vague, non-procedural questions as sensemaking (Gotz, 2007; Russell et al., 1993). Although there is some confusion regarding what exactly constitutes sensemaking, I will use the above definition throughout the rest of the dissertation. In the sensemaking process, individuals do not rely on a single source of information. Rather, they integrate multiple sources of information and synthesize that information together.

In this chapter, I will provide the rationale of why I chose this definition of sensemaking to use in my dissertation. I believe that within the context of the expertise seeking behavior I wish to elucidate, my research goals are best served by using this definition. But before that, let me show different examples of sensemaking. By comparing their commonalities and differences, these examples help us gain some idea of what sensemaking is. These examples have been adopted from Qu (2006).

#### *Sensemaking example 1*

Tom is a doctoral student who just starts working in a new research area. He wants to gain some sense of this unfamiliar field. He starts with some questions, such as “What are the basic concepts and methodologies in this field?”, “What are the popular research topics?”, “What has been done?”. He makes a foray into the literature

and starts to get a better understanding of the field, such as the categories of research sub-topics, and different research methodologies. As he reads more literature, he develops more complicated mental models for the field and accumulates more related material (books, papers, websites, notes, email, etc.)

*Sensemaking example 2*

Mary tries to figure out a more efficient layout for her kitchen. She slowly figured out problems of current layout by using the kitchen. For example, she had put spices on a particular shelf and later found it inconvenient to turn around every time to reach them when she cooks. As a result she decides to buy a rack and puts spices next to the oven.

*Sensemaking example 3*

John and Susan, a young couple who have just had a baby, need to buy some nursery furniture. They find out what they need as they go along. They start with “we only need a crib for the baby”. This initial model soon proves inadequate when they discuss and think more about their task of taking care of a baby (“where are we going to change the baby’s diaper?”) or after they gather more information (“I saw a nice rocker in a store today. Do you think our baby will need that?”). The young couple gradually learns not only more about the world, but also their own needs. At last, they realize that “we need several items of baby furniture to do different things”. In this process, the husband and wife collaborate with each other. They negotiate what to look for. They divide and conquer, coordinate, share, and co-evaluate.

*Sensemaking example 4*

A newly established academic department is trying to make sense of what its new programs should look like. Through several years, people in the department discuss the curriculum, the possibility of various research directions, and what types of new faculty should be recruited. They also go out to visit similar programs and invite

people to give talks. The identity of the school is gradually established by the efforts of various people in this department and their interactions with the outside world.

These examples share an important commonality: people face new problems or unfamiliar situations, and some type of knowledge (internal or external, individual or social) is gained in service of the task people want to do. At the same time, these examples lay out a nice range of behaviors under a very general sense of sensemaking. The difference among these examples are obvious: some of them are at the individual level (example 1 and 2), some are at the group (example 3) or organizational level (example 4); some involve more explicit knowledge (example 1), some involve more tacit knowledge (example 3); some happen only in sensemakers' heads (example 2), some involved external artifacts, settings, etc. (example 1).

With these various sensemaking examples in mind, below I will examine two sensemaking models that have been particularly influential in the information science and human computer interaction literature. These are 1) the sensemaking model by Brenda Dervin (1992), and 2) the sensemaking model by Dan Russell and his colleagues (Russell et al., 1993). Although there are many models and views of sensemaking (G. Klein et al., 2006a), these models provide a useful point of departure in explaining the information analysis and synthesis skills individuals undergo when looking for experts. It is noteworthy that Karl Weick's (1995) model of sensemaking is an influential model of sensemaking. However, it is meant to be applied at the organizational level and is thus beyond the scope of this dissertation.

A general sensemaking model is proposed by Dervin (1992), as shown in Figure 3. In this figure, "Situation" refers to the time-space context where sense is constructed. "Gap" is the disparity between user's current knowledge and the knowledge needed to accomplish the task. It is also known as the information need. People bridge the gap when they construct sense and move through the time-space

context. “Help” can be regarded as outcomes of sensemaking that help to bridge the Gap. Dervin pointed out that sensemaking is a cyclic process in which a sensemaker starts in some situation where he needs to make sense of something. This information need drives him to seek help. After receiving the help (sense is made), he is in a new situation with new gaps that need to be bridged.



Figure 3. Dervin's sensemaking triangle: Situation-Gap-Help

Dervin's model gives us a highly abstract framework of sensemaking processes conducted by individuals. There are several important points made by her model.

First, knowledge is emphasized as one of the central concepts of sensemaking. Dervin defined sensemaking as the process of detecting and filling the knowledge gap. Her theory mainly focuses on how to help people catch/express/communicate the knowledge gap. A “Sense-Making approach” is suggested in interviews to help respondents describe the situation and their question or confusion in that situation (Dervin, 1992; Dervin & Dewdney, 1986).

Second, the whole sensemaking process is posited in a situation. The situation is the time/space context where the sensemaking problem (knowledge gap) arises, where the sensemaker gathers information to solve the puzzle, and where the sensemaking results are evaluated and the actions are taken. The situation/context

greatly affects the sensemaker's behaviors and decisions. Some aspects of the situation/context, such as the available information resources, the importance of the task, etc., influence the cost structure underlying the sensemaking process, thus influence sensemaker's behavior. Some aspects of the situation/context, such as the status of the world at the time when the sensemaking problem occurs, give more information about the thing or event people are trying to make sense of so that people could make better sense of it. However, the importance of context puts many challenges on how computer systems could support sensemaking. Is it possible for computer systems to capture important aspects of a situation? How much of the context information can be caught, and at what cost? How could such context information be used automatically in sensemaking?

Third, an important aspect of Dervin's theory is the coupling of sensemaking and information seeking. The knowledge gap is the origin of information need, thus, to describe the knowledge gap is to express the information need. A sensemaker seeks information in order to fill the knowledge gap. Serving as both a sensemaking model and an information seeking model, Dervin's theory reveals the tightly coupled relationship between sensemaking and information seeking: sensemaking is the incentive or the ultimate goal of information seeking. Information seeking is one link in the iterative cycle of sensemaking. The implication from such a relationship is that for a sensemaking supporting system, the information seeking process needs to be supported in the sensemaking context.

Despite all the valuable points brought by this model, its abstractness and generality decreases the analytic and computational power of the model. First, too many activities could be considered sensemaking under this general definition. There is a need to further categorize the variety to allow an investigation of various features and characteristics of different sensemaking activities. Second, this general model



does not reveal different activities and steps in the process of sensemaking (except a general information seeking idea). This hinders the effort to analyze and assist sensemaking activities. This abstract, cognitive-behavior model does not offer many implications about how computer systems could be involved to help the sensemaking process.

In summary, Dervin's sensemaking model contains many interesting insights into sensemaking, including the importance of the knowledge gap and knowledge acquisition, the context of sensemaking, and the relationship between information seeking and sensemaking. However, the generality of the model does not help a closer investigation of the sensemaking process people undergo while looking for experts.

Russell et al. (1993) proposed a more specific model of sensemaking, which posits the use of representations in service of accomplishing a task. They define sensemaking as "a process of searching for a representation and encoding data in that representation to answer task-specific questions". Figure 4 shows the representation development in a sensemaking process. A sensemaker starts with an initial representation which he thinks could capture salient features of the information in a way that support the accomplishment of the task (the generation loop). Then he identifies information of interest and encodes it in the representation (the data coverage loop). However, when the sensemaker's understanding of the sensemaking task grows, he may find that the initial representation is not adequate to characterize the sensemaking problem, which may impair the accomplishment of the sensemaking task. When this mismatch between his representation and the task (called "residue") becomes sufficiently problematic or costly (in terms of effort), the person is increasingly motivated to find a better representation, intending to reduce the cost of task operations (the representational shift loop). The new representation is then used

for encoding information, until sufficient residue builds up and yet a better representation is needed or the task can finally be satisfactorily accomplished.

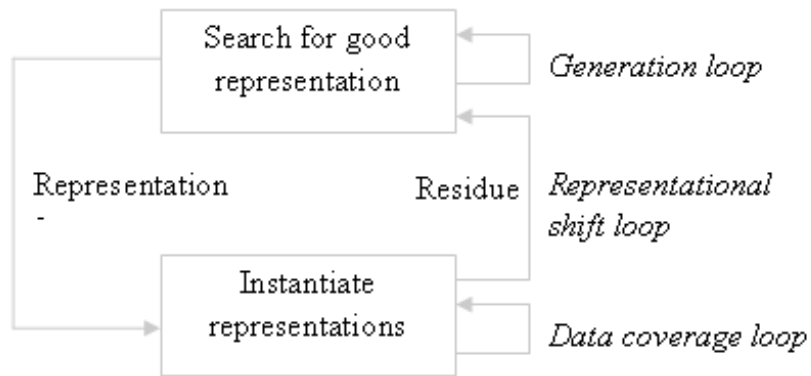


Figure 4. Representation development in Russell et al.'s sensemaking model

In Russell et al's model, a sensemaking process contains a cycle of representation search, information encoding, evaluation and representation shift. The decomposition posits a framework of different activities involved in sensemaking, enabling closer investigation. There are several traits of this framework that merit more discussion.

First, the search for representation schema is separated from the encoding of information into the representation. This is essentially the separation between structure and content. This separation simplified the problem by focusing on two processes, one emphasizing the structure construction, one emphasizing data collection. We can think about cases where these two processes are quite independent. For example, when you are shopping for a digital camera online, you may decide to use a table representation to compare features of different models before actually seeking data that fit in the table structure. This separation allows us to study the representation construction and

the data collection activities individually and allow us to focus on the hard part of the problem - the representation shift and construction.

However, this separation may oversimplify the sensemaking process. In some cases, it is hard to separate the search of representation schema from the search for encoding information. First, structure and content often co-exist in nature representations and can be hard to separate. For example, when a person finds a list of digital cameras, she finds both content and a simple structure over the content (the list). Second, the structure and content may grow simultaneously in a sensemaking process. For example, people's file folders are often built along with the growth of their files, making subfolders when the content of the folder is large and therefore hard to peruse.

Second, Russell et al. introduced the concept of “residue”, which refers to the unfitness of the representation to the sensemaking task. The concept of residue, together with the cost structure, explains the incentive for representation shift, because residue may make the execution of the task costly. Sometimes, people have concrete ideas about residue, such as data that cannot be encoded or is missing in the structure, or the unusable part of representations. However, in many cases, it's hard to identify and explicitly express the unsuitability of a representation for a task. Sometime people even have no idea what is wrong with the current representation or whether there will be a better representation. If we consider residue as the difference between the current representation and a better one, there is a wide range of possibilities to explore, such as the problem space, the type of representation, the appropriate granularity or scale. Therefore, how to detect the residue and help people to reduce the residue is a big challenge to sensemaking researchers. This is one of the areas that could be supported with technology.

Third, Russell et al. did not talk about situation, context, and action in sensemaking as explicitly as in Dervin's theory. Instead, they talked in detail about one related factor that shapes the sensemaking process - the cost structure of the task, which depends on the abstract problem of the task, the physical setting of the task, the background knowledge of the sensemaker, the available information resources, among others. In Russell et al.'s model, the cost analysis together with the process decomposition explains the motivation for representation seeking and evolution in sensemaking. A sensemaker shifts her representation when the anticipated change is expected to bring more benefit than the expected cost of change. The cost analysis is also a systematic approach to diagnose sensemaking tasks and locate the costly parts where some improvements could be made. For example, in the laser printer case Russell et al. studied, the cost analysis shows that the most time-consuming activity in that sensemaking task is data extraction – “finding the relevant documents, selecting the information, and transforming the information into canonical form”. Therefore, a reasonable suggestion is adopting some automatic information processing tools to shorten the time spent on data extraction.

Both Dervin's and Russell et al.'s models take an information-centered or knowledge-centered view of sensemaking. But Russell et al.'s model has a more concrete framework with a more narrowed focus. In the description of these two models, we can see their similarity, with Dervin's “building a bridge” approximately corresponding to Russell et al.'s “constructing a representation”. Taking a very general representationist stance, where we consider knowledge representation to be both implicit and explicit, internal and external, the knowledge gaining process is the process of knowledge representation evolution. Then even Russell et al.'s sensemaking model could be regarded as a process of gaining knowledge to bridge the knowledge gap.

However, the coverage of the two models is different. Dervin's sensemaking model could cover all categories of knowledge representation by generally talking about "Knowledge." On the contrary, Russell et al.'s model only covers explicit representation and their case study had a focus on explicit, external representations, which is the only category of knowledge representations that is directly accessible by outside observers or computer systems. They urged exploring the potential of computer manipulations of explicit external representations to enhance sensemaking.

In summary, Russell et al.'s sensemaking model brings representation into the center of the sensemaking study. Its decomposition of the sensemaking process enables close investigations of different parts of this process. Compared with Dervin's model, it is more specific and more narrowly-focused.

While Dervin and Russell et al.'s sensemaking models are the primary impetus for this dissertation, let me briefly discuss some of the other sensemaking models in the literature. Sensemaking has long been studied in sociology. A survey of the development of organizational sensemaking theories could be found in Weick's *Sensemaking in Organizations* (Weick, 1995). Different studies revealed different aspects of sensemaking: Starbuck and Milliken (1988) and Westley (1990) pointed out that sensemaking involves placing stimuli into some kind of framework; Louis (1980) viewed sensemaking as a process in which people cope with interruption and use retrospection to explain surprises; Thomas, Clark and Gioia (1993) mentioned the reciprocal interaction of information seeking, meaning ascription, and action in sensemaking processes.

Weick (1995) gave a more comprehensive definition of sensemaking through seven properties: grounded in identity construction, retrospective, enactive of sensible environments, social, ongoing, focused on, and by extracted cues, driven by plausibility rather than accuracy. He gave a rich description of sensemaking at both

individual level and organizational level. At the individual level, other than taking the representationist stance and focusing on the change of representation in a sensemaking process, he showed important properties on how people gain the sense (i.e. the knowledge in Dervin and Russell et al.'s model), such as the retrospective, ongoing, etc. More importantly, he showed how people make sense in an organizational environment through the interactions with the social system, such as identity construction, etc.

The sensemaking claims suggested by Dervin and Russell et al. provide a rich point of departure for my dissertation. Inherent in all these models is placing stimuli into some kind of frame that allows an individual to construct meaning by comprehending, understanding, explaining, attributing, extrapolating, and predicting. When new stimuli fit existing frames and expectations the sensemaking process goes unnoticed. When faced with complex, uncertain, and non procedural tasks, it is rare that new stimuli will fit into existing frames. When stimuli do not fit a frame, uncertainty emerges and that is when sensemaking requires conscious and social interpretation of the discrepancies. Lipshitz, Klein, Orasanu, and Salas (2001) propose that uncertainty is “the sense of doubt that blocks or delays action” (p. 37). Such doubt can arise from a variety of sources. Information may be missing or too complex to make sense of. Time constraints and high stakes may induce second guessing. Implications and consequences may be unknown.

We see strong parallels in this situation with that of searching for experts using technology. As mentioned previously, seeking to contact others for expertise using technology involves a set of interconnected cognitive activities, including generating a query, searching for relevant information, evaluating and making sense of information found, and coherently integrating different pieces information into a coherent whole to

arrive at a decision. It may involve sifting through massive volumes of information under deadline pressure to make complex search decisions under uncertain conditions.

I will thus define the process of gathering complex, changing and potentially equivocal information, and comprehending it by *connecting nuggets of information* from *many sources* to answer vague, *non-procedural* questions as sensemaking (Gotz, 2007). This definition suits this research best since I am uniquely positioned to research all the aspects of sensemaking outlined in this definition. Chapter 3 will describe in detail how this dissertation research addresses the various elements of sensemaking mentioned in the above definition. Moreover, many of the sensemaking claims have yet to be tested empirically through field-based studies. Therefore my research directly contributes to this body of literature. Furthermore, I believe that the process through which individuals synthesize information about a person available through various digital artifacts into a coherent whole is a form of ‘people sensemaking’. Existing research on sensemaking has investigated sensemaking from maps (Bauer, 2002), web documents (Gotz, 2007; Qu, 2003; Qu & Furnas, 2005), medical question-answering tasks (Billman & Bier, 2007), hand-off of tasks (Sharma, 2007), and front-end project and technology selection (Bergman & Mark, 2002). To the best of my knowledge, there are very few, if any, studies on the sensemaking process when looking for people using information technology. An aim of this dissertation is to thus disambiguate the processes surrounding ‘people sensemaking’.

In the next section, I will discuss how concepts from existing theories of information search are relevant within the context of expertise search.

### **Theories of Information Search**

When talking about search in the context of information retrieval, it is imperative to discuss existing theories of information search. I will discuss two such theories or models: 1) information foraging, and 2) berrypicking.

Information foraging behaviors underlines the way people look for information. This theory (Pirolli & Card, 1999) provides a way to understand how people search for usable information. Humans follow certain patterns of behavior by virtue of being a member of the animal kingdom. These patterns can be observed on a very fundamental level, underlying the more apparent ‘taught and learned’ behaviors. These patterns manifest in animals as essential elements for survival in an environment abundant with resources such as food, but with a cost associated with each resource. Due to a variable cost associated with each resource, there exists an optimal mechanism to maximize the resources gained per unit of the associated cost. Humans seem to follow the same patterns looking for information as animals do foraging for food. It is observed that humans apply similar optimizing behaviors while foraging for information. Information too, like food, can be considered to be an available resource with an associated cost of consumption. Information foraging theory attempts to explain such an information seeking behavior in humans.

According to this theory, humans follow in-built behavior patterns to minimize the effort required in seeking information. Hence their information seeking endeavors always tend to converge to optimized search paths. Ideas from optimal foraging theory are applied in the context of information to arrive at the results found in Information foraging theory. These foraging behaviors have evolved over many years in animals. They have developed in-built mechanisms that naturally tend to maximize the amount of food obtained per unit of effort.

This analogy is very well explained in the following extract from Pirolli and Card (1999) – “Imagine a predator, such as a bird of prey, that faces the recurrent problem of deciding what to eat, and we assume that its fitness, in terms of reproductive success, is dependent on energy intake. Energy flows into the environment and comes to be stored in different forms. For the bird of prey, different



types of habitat and prey will yield different amounts of net energy (energetic profitability) if included in the diet. Furthermore, the different food-source types will have different distributions over the environment. For the bird of prey, this means that the different habitats or prey will have different access or navigation costs. Different species of birds of prey might be compared on their ability to extract energy from the environment. Birds are better adapted if they have evolved strategies that better solve the problem of maximizing the amount of energy returned per amount of effort. Conceptually, the optimal forager finds the best solution to the problem of maximizing the rate of net energy returned per effort expended, given the constraints of the environment in which it lives.”

A few key concepts have emerged out of this theory. Information can be considered to be a resource which has a cost associated with it, similar to the cost associated with obtaining food. Hence there exists a combination where the amount of resources obtained can be maximized per unit of effort. Information availability is patchy in nature. Distribution of information is not continuous, but is clustered in patches. Hence information can be available in patches and effort is required both to find information inside the patches and to traverse between information patches. The values of information can be gauged by metadata and other proximal clues. This determination of value is called ‘information scent’ (Pirulli & Card, 1999). Information foragers use this idea to seek out the desired information and naturally then to assume that the information with a stronger information scent has more value than the one with a weak information scent. According to the availability of the information foragers tend to follow an information diet. They give preference to one source of information over the other to obtain maximum amount of information with minimum effort.

Ideas from this theory will be used in my dissertation to discuss how ‘information scent’ acts as a ‘signal’ when individuals search for people online. These signals provide seeking relevant experts easy and with minimum effort. While animals rely on scents to indicate the chances of finding prey in current area and guide them to other promising patches, humans rely on various signals in the web information environment to select the most promising sources of information. Transitional behavior is also observed in animals when they seek prey for food. Animals will move from one food patch to another food patch to catch the prey. A similar analogy can be drawn about humans searching for people online. Each search result set can be considered to be a patch of information. Humans can obtain one result set, and then reformulate their query to obtain another result set.

Another model of information search is the Berrypicking (Bates, 1989) model. This model is considered to be closer to actual information searcher’s behavior and hence is much superior to the traditional information retrieval model. The Berrypicking model provides an explanation to better understand the complex task of an information searcher.

The Berrypicking model departs from the traditional information retrieval model in four major areas, namely the query formulation, the search process, types of techniques used for searching and in terms of the search domain. The classic model of Information Retrieval is based on the fundamental idea representing the user’s information search. According to this theory the user presents a single query, which is then matched to contents of a dataset to yield just one output set. Hence this theory presents information search as a single step process from the query to the end result. Even though this provides a conceptual understanding of a simple search process, it is not adequate to model the more complex information searches. In actual searches observed in real-life, the user seldom starts with more than one set of requirements.

The searcher usually begins with just one set of requirements or just a single reference and then fans out the search in newer directions after coming across other relevant sources. Usually the user starts with one query. Each relevant result gives the searcher newer ideas for subsequent search queries. This continuous process of modifying the query at each stage of the search to get better results can be considered to be forming a query-result feedback loop. The query formulation itself undergoes a change at each stage and can be considered to be continuously evolving. Hence this type of search is called an evolving search. Also, the query results obtained at each stage contribute bit and pieces to the complete result set of the searcher. Hence the final results can be considered to be a collection of bits of information retrieved at each stage in an ever evolving search.

An analogy can be seen in picking berries on bushes. The berries are scattered all over the bushes and do not come in bunches. These berries have to be picked one at a time. Similarly users usually start with just one relevant reference and move through a variety of sources, each new piece of search result providing a new conception of the query. At each stage a user modifies both the query terms as well as the search requirements. This type of ever changing search is called an evolving search. Some salient features of this theory are dynamic nature of the query, final information as a collection of the results of an evolving search, and use of a variety of search techniques and sources to obtain the search results.

In summary, the Berrypicking model is a model for searching online and can be considered to be closer to the real behavior of information searchers than the traditional model of Information Retrieval. The nature of this model is similar to the nature of Berrypicking and can be used to improve online interface designs. The salient points of the Berrypicking model are as follows -

- Multi stage search with feedback.

- Evolving search.
- Better represents user's search compared to traditional Information Retrieval.
- Query can change based on the previous results.
- Query is satisfied by collecting bits and pieces of information from the results.
- Information seeker zigzags through information space and varies search strategies to reach results.

The Berrypicking model is extremely helpful for understanding search behavior. However, the nature of the tasks the participants in my studies performed were somewhat artificially constrained to limit a wide range of behaviors in order to ensure experimental comparability. I will discuss these behaviors through ideas from the Berrypicking model, but their application will be limited.

In the following section, I will discuss relevant concepts from the social network analysis literature and how they may apply to expertise search. Recently, researchers have realized the importance of various social network characteristics on how people select whom to contact for expertise. A review of these social network analysis concepts is thus relevant.

### **Expertise searching and social networks**

A social network is the infrastructure for interpersonal information interactions. Its structure and dynamics heavily influence people's expertise seeking processes. Researchers in expertise sharing have recently started to note the importance of social networks and built systems using social networks as channels for expertise sharing (Kautz et al., 1997a; McDonald & Ackerman, 2000). Currently, there are basically two lines of social network research: research in the field of sociology and research in the field of statistical physics. Each field has a different

research focus and uses different methods.

In the field of sociology, social network analysis (SNA) focuses on relationships between actors rather than attributes of actors (Wasserman & Faust, 1994). Based on the mathematical foundations of graph theory, statistical and probability theory, and algebraic models, SNA provides a set of metrics to study network properties, at the following levels.

- Individual actor level: connectedness, reachability, prominence, betweenness, isolation, and centrality.
- Dyads, triads, and group levels: reciprocity, symmetry, transitivity, clustering coefficient, and cohesions.
- Global level: network density, connectivity, heterogeneity.

In the field of statistical physics, research has focused on common properties of many different kinds networks, including social and non-social networks (i.e. Internet, World Wide Web, and biological networks). The research topics include topology, evolution, and complex processes occurring in networks (Dorogovtsev & Mendes, 2002; Newman & Park, 2003). Compared to focusing on various metrics that measure the individual or network attributes in the field of sociology, research in this area usually focuses on the general scaling properties of the network, such as the so-called “scale free network” and “small world effect”. Findings in this area have given computer science researchers great help in designing better searching algorithms in various information networks such as the web, p2p file sharing, and blogs (Adamic, 1999; Adamic, Lukose, Puniyani, & Huberman, 2001; Adar & Adamic, 2005; Brin & Page, 1998; Menczer, 2002).

For the purpose of this dissertation, I will focus only on several topics I feel are important for expertise searching research. In next two sub sections, I will first survey

related work on the searchability of social networks, as well as how can we search them efficiently. Then, I will look at some social network characteristics that are important for information searching.

### ***Small world***

The classic study on searching in social networks is the “small world” experiment. In the late 1960’s, Milgram and Travers found that subjects could successfully send a small packet (with a name, the city, and the profession of the recipient on it) from Nebraska to people in Boston (Travers & Milgram, 1969). The subjects did so, even though they had only local knowledge of their acquaintances, by passing the packet to an acquaintance that they believed to be closest to the target. Travers and Milgram found the average length of acquaintance chain is roughly six. The result of this experiment indicated that the social network is searchable and that the paths linking people are short, which is often referred to the “six degrees of separation” phenomena.

A key question in such experiments is how people select the next person to forward the packet or message from among hundreds of acquaintances, which ultimately leads to a short chain between the sender and the target. Later experiments found that geographic proximity and similarity of profession to the target are the most frequently used criteria by participants (Bernard, Killworth, & McCarty, 1982; Dodds, Muhamad, & Watts, 2003; Killworth & Bernard, 1978). For instance, in Dodds et al.’s global level small world experiment that involved 60,000 email users and 18 target persons in 13 countries, they found that the geography proximity of the acquaintance to the target dominated the early stage of the chain, because senders are geographically distant. Occupational proximity was used more frequently after the third step. Other related findings in Dodd et al.’s experiment is that successful searches were conducted primarily through intermediate to weak strength ties, and that

the success of the search did not rely on a small minority of exceptional individuals (i.e. social hubs).

Recently, mathematical models have been proposed to explain why these simple heuristics are good at forming short paths (Kleinberg, 2000; Watts, Dodds, & Newman, 2002). In general, I prefer the hierarchical network model of Watts et al to Kleinberg's. It assumes that the social network usually has a structure, in which individuals are grouped together by occupation, location, interest, and so on. As well, these groups are grouped together into bigger groups and so forth. The difference in people's group identities defines their social distance. By choosing individuals who have the shortest social distance to the target at each step, people can gradually reach the target in a short path with only local information about their own immediate acquaintances.

The analysis above is preliminary. However, we can see that there are many similarities between searching a named person and searching any person that carries wanted expertise. Building a similar small world model for expertise searching would be a very interesting research topic.

### ***Automatization of network searching***

In those small world experiments, it is a person who decided to whom the messages were forwarded. Since participants knew the target's location or profession as well as their own local neighbors' related attributes, with the help of their own understanding of the relations and similarities between the target's and their neighbor's identifiable characteristics, they could pick the next person in the searching chain effectively.

Adamic and her colleagues did several simulation studies to explore strategies that could be used in the automatization of the network searching (Adamic & Adar, 2005; Adamic et al., 2001). They found that the best-connected searching algorithm

that makes use of the skewed degree distribution of many networks is an efficient algorithm in power law networks. By passing the query to highly collected nodes first, the query can be spread broadly in the network and find the desired results quickly.

Similar algorithms were later adopted in peer-to-peer file sharing networks, such as Gnutella, to replace the traditional broadcast strategies. Compared to the classical breadth-first-search algorithm, which can find the target quickly but with extremely high cost in terms of bandwidth, searching utilizing these high degree nodes proved to be relatively fast and used much less resources.

In another computer simulation study on the HP email network, Adamic and Adar (2005) found that some simple strategies are more effective than best-connected strategies in automatically finding a named person with some known identities, such as using a contact's position in physical space or an organizational hierarchy. Adamic and Adar suggested that this was due in large part to the agreement with theoretical predictions by Watts et al. and Kleinberg about optimal linking probabilities relative to physical space or in the organizational hierarchy.

In summary, Adamic's studies suggest we can find efficient ways to automatically navigate to a person in social networks. Then, is it possible to use similar approaches to automatically search for expertise in social networks?

### ***Automatization of expertise searching in social networks***

Recently, some work has been done on automating expertise searching in social networks (Yu & Singh, 2003; Zhang & Ackerman, 2005). It is different from the work of Adamic and her colleagues or other small world experiments in which the desired person is known by name or unique identifier. In the expertise searching problem, a suitable person or set of people is not known in advance. One must be found by matching people against a list of attributes.



In their work on “MARS” referral system, Yu and Singh (2003) proposed a distributed expertise searching algorithm and studied related dynamics using simulation. They used the similarity between a query vector and a neighbor’s expertise vector, plus some consideration of one’s historical referring performance, as the criteria for picking the next agent in a referral graph. The simulation results using a scientific co-authorship network indicate using “information scent” can help people find experts in such a network.

Following Adamic et al and Yu and Singh’s work, Zhang and Ackerman (2005) compared various strategies that could be used in searching expertise in social networks. They found that using highly connected people or using weak ties is more efficient regarding the searching speed and per-query cost than other strategies. More importantly, they found that the “information scent” strategy is not as efficient as Yu and Singh (2003) claimed.

There could be many reasons for these different results. First, Yu and Singh (2003) never compared their searching strategies with other possible strategies. Second, Yu and Singh’s (2003) simulation was conducted in a co-authorship network while Zhang and Ackerman’s (2005) simulation was on an email network. Information distribution on these two types of networks may be different. These results and discussions suggest that we should further look at how information is distributed in social networks.

### ***Important network characteristics that affect network searching***

We have discussed the searchability of social networks in previous sections. But to design better searching strategies, we need to understand what characteristics of social networks are important. In this section, we will look at three of these characteristics, including: structural properties of social networks, various centrality measures, and impact of ties.

### *General structural characteristic of social networks*

A social network is usually represented as a graph. However, different from a random graph or other non-social networks, the structure of social networks is highly meaningful and has its special characteristics.

The small world network model suggests a general characteristic of many large scale social networks. The key idea of the small world network model is that most people have a relatively small circle of friends who generally all know each other, but the shortest-path length from one person to any other in the whole world is possible very short (Newman, 2003).

Newman and Park (2003) further proposed two important properties that differ between social networks and non-social networks:

- Different patterns of correlation between the degrees of adjacent vertices: Degrees are usually positively related in most social networks while negatively correlated in most non-social networks. In other words, in social networks, a person who has a lot of social connections tends to connect to other persons who also have a lot of social connections.
- Level of clustering or transitivity: Social networks usually show a high level of clustering while non-social networks do not.

### *Centralities of actors*

The studies on the structural properties of networks have mostly been concerned about an actor's position in a network, which can affect his role in information dissemination and access. The key idea is that people in different positions in a network will have different access to information, resources, and social support. The most commonly used measures of people's network position are centralities. There are many different types of centralities (Bonacich, 1987; Freeman, 1979; Newman, 2005). Following are several widely used ones.

The simplest one is degree centrality, which simply counts the number of direct connections an actor has. In general, a person with high degree centrality is viewed as socially popular and is like a social hub. The best connect strategy used in Adamic's simulation used this type of centrality. Furthermore, there are in-degree centrality and out-degree centrality that consider the direction of social ties. A person with high in-degree is good at collecting information, while a person with high out-degree is good at spreading information. The weakness of degree centrality is that it takes into account only the immediate ties that an actor has, rather than indirect ties to all others.

To address the weakness of degree centrality, closeness centrality approaches consider the distance of an actor to all others in the network by focusing on the distance from each actor to all others instead of only to local ones. Depending on the definition of "close", there are several slightly different measures for closeness centrality, such as the ones based on the Eigenvector of geodesic distance or based on reachability. People with high closeness centrality are in an excellent position to monitor the information flow in the network, and they usually have the best visibility into what is happening in the network

Betweenness centrality is another important centrality measure of information flows in the network. It examines "the extent to which an actor is situated among others in the network, the extent to which information must pass through them to get to others, and consequently, the extent to which they are exposed to information circulation within the network" (Freeman, 1979, p. 215). If a person has high betweenness centrality, he frequently acts as a local bridge that connects the individual to other people outside a group. The technological gatekeepers mentioned in Allen's (1977) study probably had high betweenness centrality.

There are also multiple variants of betweenness centrality, such as ones based

on information flow or based on random walk. These measures provide us methods to quantitatively describe the network structure as well as to compare individuals' differences. More importantly, by comparing these different measures and noting how sociologists explain them, we can better understand that connections among people are not uniformly distributed in the social network. Unlike a theoretically constructed graph, the connections among people in a social network are highly meaningful and vary greatly (Newman, 2003; Newman & Park, 2003). People with various degrees in social networks also vary on their information access abilities as well as social status (Wasserman & Faust, 1994). People in different network positions need to be supported differently in designing peer-to-peer based expertise sharing systems because of different accessibility and workload concerns.

#### *The impact of ties*

An individual's network position affects his overall ability to access and diffuse information. However, for each individual's information seeking behavior in social networks, the strength of his social ties may have an important impact. The connections between two individuals can have different strengths. The strength of association varies and is not always symmetrical. Usually, in social networks, the strength of association is divided roughly into strong and weak ties. The term of weak tie is firstly used by (1973) to represent the ties in a social network that are not strong, such as loose acquaintances that people met at a party. By contrast, strong ties usually mean those who are kin relations or close personal friends. These different tie strengths have different benefits and tradeoffs in searching for information. Weak ties display an important bridging function, allowing information travel from one subgroup to another subgroup in a social network. They can help people get new information and adopt innovations (Brown & Reingen, 1987; Burt, 2004; Granovetter, 1973; Haythornthwaite, 2002). Strong ties have found been more likely activated for

the flow of referral information. They are usually perceived to be as bearing lower social psychological cost in the searching process (Allen, 1977; Brown & Reingen, 1987; Granovetter, 1973). When designing local searching algorithms, one needs to consider tie strength.

The findings in social networks research we discussed above can provide many aids to an expertise sharing study. They provide a deep understanding of the structure that underlies expertise sharing activities. They also provide us methods and tools to analyze this structure. More importantly, they may provide us a new ways of designing expertise sharing system searching expertise in social networks. Different from previous peer-to-peer based referral systems, such new systems should emphasize the understanding of the human social network, and small world network searching problem, as well as consider the impact of various network structure properties and the characteristics of social ties.

In the next section, I will discuss how signaling theory can be used as a rich interpretive scheme to shed light on human phenomena such as ‘people sensemaking’ while looking for expertise.

### **Signaling theory**

Signaling theory provides a useful framework in understanding which pieces of information may be more reliable when making inferences regarding a person’s expertise. Reliable signals are pieces of information that are hard to fake. Such information allows users to ‘separate the grain from the chaff’ by distinguishing between different types of information. For example, social network connection information may be a more reliable signal of expertise because people within a social network connection chain can credential the expertise of an individual.

Signaling theory is in essence a theory of communication. It describes a process of discerning and interpreting conveyed information. This theory may be

particularly useful when applied to situations in which a wide variety of information is conveyed about an individual, and one needs to determine the credibility of such information.

Signaling theory has its origins in both economics and biology. Let me use examples from the animal kingdom and further examples from economics to motivate my use of signaling theory.

In biology, signaling theory has been used to explain seemingly wasteful and detrimental ornaments and behaviors in animals (Zahavi, 1975). The signal itself, carried in behaviors and other phenotypic traits, is costly in terms of time, energy, or risk, making it difficult to fake, and ensuring that the signal transmits reliable information to the signal receivers.

Among the frequently cited examples of costly signal use in predator deterrence is stotting in gazelles (Zahavi, 1975). When a gazelle notices a predator, the gazelle stomps its feet and turns away from the predator, showing a black and white rump. Then the gazelle will stot, jumping high into the air on all four legs. Although this behavior reveals the gazelle to the predator, it also serves as a reliable signal that the gazelle is in good physical condition and is likely to outrun the predator if pursued. Because stotting requires great energetic expenditure and wasteful use of valuable “escape time,” only gazelles that are in good condition will stot. For this reason, stotting is a reliable signal to the predator that a long, difficult pursuit will only result in failure and exhaustion. Zahavi and Zahavi have identified several means by which animals signal to competitors, including singing, aerial display, electric pulses, posturing, and the release of chemicals (Zahavi & Zahavi, 1997). Physical attributes can also serve as honest signals of quality to attract potential mates. For instance, bright coloration in males is an honest signal of quality because it is likely to attract the attention of predators. These colors may also attract female attention to size,

shape, and movement of the males (Zahavi & Zahavi, 1997). Those males who are able to survive with these bright colors may be higher in quality and more desirable. Common examples include the massive tail feathers of male peacocks. These characteristics make the male more vulnerable to recognition and attack by predators, and require strong physical constitution and adequate nutrition. For these reasons, ornaments can be honest, reliable signals of quality in a mate.

When a toad and his rival vie for the same mate, each faces an important strategic decision. Should he fight for her or set off in search of another? To fight is to risk injury, but to continue searching has costs as well. At the very least, it will consume time. And there is no guarantee that the next potential mate will not herself be the object of some other toad's affections.

In deciding between these alternatives, each toad's assessment of the other's fighting capabilities plays an important role. If one's rival is considerably larger, the likelihood of prevailing will be low and the likelihood of injury high. So it will be prudent to continue searching. Otherwise, it may pay to fight.

Many of these decisions must be made at night, when it is hard to see. Toads have therefore found it expedient to rely on various non-visual clues; the most reliable is the pitch of the rival's croak. In general, the larger a toad is, the longer and thicker are its vocal cords, and hence the deeper its croak. Hearing a deep croak in the night, a toad may reasonably infer that a big toad made it. Indeed, experiments have shown that the common toad is much more likely to be intimidated by a deep croak than a high-pitched one (Krebs & Dawkins, 1984).

The above examples illustrate two important properties of signaling theory: 1) signals must be costly to fake and 2) if some individuals use signals to convey favorable information about themselves, others will be forced to reveal information even when it is considerably less favorable. Each principle is important in

understanding how information is gathered and interpreted. I will begin by stating each principle in terms of its application in the toad example and then proceed to examine its application in a wide variety of contexts.

***The costly to fake principle***

For a signal to be credible, it must be costly (or, more generally, difficult) to fake. If small toads could imitate the deep croak that is characteristic of big toads without much cost, a deep croak would no longer *be* characteristic of big toads. But they cannot. Big toads have a natural advantage, and it is that fact alone that enables deepness of croak to emerge as a reliable signal.

This costly to fake principle has clear application to signals between people. It is at work in the following episode from Joe McGinnis's *Fatal Vision* (McGinnis, 1983). Captain Jeffrey MacDonald, an Army Green Beret physician, has been told he is suspected of having killed his wife and daughters. The Army has assigned him a military defense attorney. Meanwhile, however, MacDonald's mother recruits Bernard Segal, a renowned private attorney from Philadelphia to defend her son. When Segal calls McDonald in Fort Bragg, NC, to introduce himself, his first question is about McDonald's Army attorney:

*"Are his shoes shined?"*

*"What?!" MacDonald sounded incredulous. Here he was, all but accused of having murdered his own wife and children, and in his very first conversation with the Philadelphia lawyer who presumably had been hired to set things right, the first question the lawyer asks is about the condition of the other lawyer's shoes.*

*Segal repeated the question. "And this time," he said later, "I could almost hear Jeff smiling over the phone. That was when I first knew I had a*



*client who was not only intelligent but who caught on very quickly. He said, no, as a matter of fact, the lawyer's shoes were kind of scruffy. I said, 'Okay in that case, trust him. Cooperate with him until I can get down there myself.' The point being, you see, that if an Army lawyer keeps his shoes shined, it means he's trying to impress the system. And if he was trying to impress the system in that situation – the system being one which had already declared a vested interest in seeing his client convicted by public announcement of suspicion – then he wasn't going to do Jeff any good. The unshined shoes meant maybe he cared more about being a lawyer."*

The condition of the attorney's shoes was obviously not a perfect indication of his priorities in life. Yet they did provide at least *some* reason to suspect that he was not just an Army lackey. Any attorney who wore scruffy shoes merely to convey the impression that he was not looking to get ahead in the Army actually *wouldn't* get ahead. So the only people who can safely send such a signal are those who really do care more about their roles as attorneys.

Below are some applications of the costly-to-fake principle:

#### *Product quality assurance*

Many products are so complex that consumers cannot inspect their quality directly. In such cases, firms that offer high quality need some means of communicating this fact to potential buyers. Otherwise, they will not be able to charge high enough prices to cover their added costs.

One way to solve this problem is for the firm to develop a reputation for delivering high quality (B. Klein & Leffler, 1981). But conditions will not always allow a firm to do this. Consider the case of sidewalk vendors that sell wristwatches on the streets of large cities. If such a 'firm' decides to go out of business, it can do so

with virtually no losses. It has no headquarters, no costly capital equipment, no loyal customers to worry about – indeed no sunk costs of any kind. Even if a vendor had supplied quality products on the same street corner for years, that would provide no assurance that he would still be in business tomorrow. And if he were planning to go out of business, his incentive would be to sell the lowest quality merchandise he could pass off. In short, a firm with no obvious stake in the future has an inherently difficult time persuading potential customers it will make good on its promises.

The incentives are different for a firm with extensive sunk costs. If such a firm goes out of business, it loses the value of substantial investments that cannot be liquidated. Accordingly, the material interests of these firms favor doing everything they can to remain in business. And if buyers know that, they can place much greater trust in the promise of a high-quality product. If such a firm charged a price commensurate with high quality and then delivered shoddy merchandise, it would get too little repeat business to survive, and would thus have incurred its sunk costs in vain.

These observations suggest a reason for believing that heavily advertised products will in fact turn out to have higher quality, just as their slogans proclaim. An extensive national advertising campaign is a sunk cost, its value lost forever if the firm goes out of business. Having made such an investment, the firm then has every incentive to deliver. That firms believe many consumers have spotted this pattern is evidenced by the fact that they often say “...as seen on national TV...” in their magazine ads.

#### *Choosing a trustworthy employee*

In many situations employees have an opportunity to cheat their employers. Many productive activities would have to be abandoned if firms were unable to hire employees who would not cheat in these situations. The firm needs a signal that

identifies a prospective employee as trustworthy. One basis for such a signal might be the relationship between a person's character and the costs or benefits of membership in specific groups. For example, perhaps trustworthy people generally enjoy working in volunteer charitable organizations, which untrustworthy people instead tend to consider highly burdensome. In such cases, the groups people decide to join may convey reliable information about their character.

This notion seems borne out in the practice whereby many professional couples in New York City recruit governesses for their children (Frank, 2001). The care of children is one of those tasks in which trustworthiness is of obvious importance since it is difficult to monitor the caretaker's performance directly. The very reason for needing someone else to look after them, after all, is that you are not there to do so yourself. Bitter experience has apparently persuaded many New Yorkers that the local labor market is not a good place to recruit people who perform reliably without supervision.

The solution many of these couples have adopted is to advertise for governesses in Salt Lake City newspapers. The couples have discovered that persons raised in the Mormon tradition are trustworthy to the degree that the average New Yorker is not. The signal works because someone who merely wanted to *appear* trustworthy would find it unpalatable, if not impossible, to have remained in the Mormon tradition. The tradition involves continuing, intensive moral indoctrination, an experience most purely opportunistic persons would find too taxing to endure. Like the deepness of a toad's croak as a signal of its size, membership in the Mormon tradition is a good signal of trustworthiness because it would be so costly for an opportunistic person to simulate (Frank, 2001). All else being equal, the perception of trustworthiness of someone belonging to the Mormon tradition is better than the average New Yorker.

### *Choosing a hard-working, smart employee*

As a final illustration of the costly to fake principle, consider a degree with honors from an elite university. Employers are looking for people who are smart and willing to work hard. There are obviously a great many people in the world who have both these traits yet do not have an elite degree. Even so, employers are reasonably safe in assuming that a person who has such a degree is both smart and hard-working, for it is not obvious how anyone without that combination of traits could go about getting an elite degree with honors.

No one really questions the fact that the graduates of elite institutions generally turn out to be productive employees. But here is a lively debate indeed about the extent to which attendance at these institutions actually causes high productivity. People who think it does point to the fact that the graduates of elite institutions earn significantly higher salaries. Skeptics caution, however, that the entire differential cannot be attributed to the quality of their education. The problem is that the students at the best institutions were undoubtedly more productive to begin with. These institutions, after all, screen their applicants carefully and accept only those with the strongest records of achievement.

### ***The full-disclosure principle***

A second important principle illustrated by the toad example can be called the full-disclosure principle, which is that individuals must disclose even unfavorable qualities about themselves, lest their silence be taken to mean that they have something even worse to hide. If some individuals stand to benefit by revealing a favorable value of some trait, others will be forced to disclose their less favorable values. This principle helps answer the initially puzzling question of why the smaller toads bother to croak at all (Krebs & Dawkins, 1984). By croaking, they tell other toads how small they are. Why not just remain silent and let them wonder?

Suppose all toads with croaks pitched higher than some threshold did, in fact, remain silent. Imagine an index from 0 to 10 that measures the pitch of a toad's croak, with 10 being the highest and 0 being the lowest. Let us suppose, arbitrarily, that toads with an index value above 6 kept quiet.

It is easy to see why any such pattern would be inherently unstable. Consider a toad with an index of 6.1, just above the cutoff. If he remains silent, what will other toads think? From experience, they will know that *because* he is silent, his croak must be pitched higher than 6. But how much higher?

Lacking information about this particular toad, they cannot say exactly. It generally will be possible, however, to make a statistical guess. Suppose toads were uniformly scattered along the pitch scale. This means that if I picked a toad at random from the entire population of toads, the pitch of its croak would be equally likely to take any value along the pitch scale. With the croaking threshold at 6, however, a toad who remained silent would be systematically different from a randomly selected toad. In particular, experience would tell that the average index for toads who remain silent is 8 (halfway between 6 and 10). Any toad with an index less than 8 would, by the fact of his silence, create the impression that he is smaller than he really is. The toad with an index of 6.1 would therefore do far better to croak than not.

Thus, if the threshold for remaining silent were 6, it would pay all toads with an index less than 8 to croak. If they do, of course, the threshold will not remain at 6. It will shift to 8. But a threshold of 8 will not be stable either. With the cutoff at that level, it will pay all toads with an index less than 9 to croak. *Any* threshold less than 10 is for similar reasons destined to unravel. This process happens not because the small toads want to call attention to their smallness by croaking. Rather, they are forced to do so in order to keep from appearing smaller than they really are.

The full disclosure principle derives from the fact that individuals do not all have access to the same information. In the toad case, the asymmetry is that the silent toad knows exactly how big he is, while his rival can make only an informed guess. As the following examples demonstrate, similar asymmetries give rise to important signals between communicators.

#### *Product warranties*

Information asymmetries help explain, for example, why the producer of a low-quality product might disclose that fact by offering only very limited warranty coverage. The asymmetry here is that producers know much more than consumers about how good their products are. The firm that knows it has the best product has a strong incentive to disclose that information to consumers. A credible means of doing so is to provide a liberal guarantee against product defects. This is credible because of the costly-to-fake principle – a low quality product would break down frequently, making it too costly to offer a liberal guarantee.

Once this product appears with its liberal guarantee, consumers immediately know more than before, not only about *its* quality, but about the quality of all remaining products as well. In particular, they know that the ones without guarantees cannot be of the highest quality. Lacking any other information about an unguaranteed product, a prudent consumer would estimate its quality as the average level for such products. But this means consumers will underestimate the quality of those products that are just slightly inferior to the best product.

Consider the situation confronting the producer of the second-best product. If it continues to offer no guarantee, consumers will think its product is worse than it really is. Accordingly, this producer will do better to offer a guarantee of its own. But because of its product's slightly lower quality, the terms of its guarantee cannot be quite so liberal as those for the best product.

With the second best product now guaranteed, the class of remaining unguaranteed products is of still lower average quality than before. The unraveling process is set in motion, and in the end, all producers must either offer guarantees or live with the knowledge that consumers rank their products lowest in quality. The terms of the guarantees will in general be less liberal the lower a product's quality. Producers clearly do not want to announce their low quality levels by offering stingy warranty coverage. Their problem is that failure to do so would make consumers peg their quality levels even lower than they really are.

When Chrysler declares, "We back them better because we build them better", we cannot be 100 percent sure it is telling the literal truth. But if the claim were grossly misleading – that is, if Chrysler cars were significantly more likely to break down than others – it would be a costly lie indeed. And therein lies a rational motive for consumers to credit Chrysler's statement.

#### *The lemons principle*

The full disclosure principle helps resolve the long-standing paradox of why new cars usually lose a large fraction of their market value the moment they are driven from the showroom. How is it, exactly, that a new car purchased for \$15000 on Wednesday could command a price of only \$12000 in the used car market on Thursday? Clearly the car does not lose 20 percent of its value within 24 hours merely because of physical depreciation.

Economists struggled for years to make sense out of this situation. In an uncomfortable departure from their characteristic professional posture, some even speculated that consumers held irrational prejudices against used cars. George Akerlof, however, suggested that mysterious superstitions might not be necessary. He offers the following ingenuous alternative explanation (Akerlof, 1970).

Akerlof began with the assumption that new cars are, roughly speaking, of two basic types: good ones and ‘lemons’. The two types look alike. But the owner of any given car knows from experience which type of car hers is. Since prospective buyers cannot tell which type is which, good cars and lemons must sell for the same price. We are tempted to think the common price will be a weighted average of the respective values of the two types, with the weights being the proportions accounted for by each type. In the new car market, in fact, this intuition proves roughly correct.

In the used car market, however, things work out differently. Since good cars are worth more to their owners than lemons are to theirs, a much larger fraction of the lemons finds its way quickly into the used car market. As used car buyers notice the pattern, the price of used cars begins to fall. This fall in price then reinforces the original tendency of owners of good cars not to sell. In extreme cases, the *only* used cars for sale will be lemons.

Akerlof’s insight was to realize that the mere fact that a car was for sale constituted important information about its quality. This is not to say that having a lemon is the only reason that prompts people to sell their cars. Even if it were just a minor reason, however, it would still keep the owner of a good car from getting full value for it in the secondhand market. And that may be all this is needed to initiate the by now familiar unraveling process. Indeed, trouble-free cars rarely find their way into the used car market except as a result of heavy pressure from external circumstances (e.g. “Going overseas, must sell my Volvo station wagon” or “Injured hand, must sell stick shift.”)

Akerlof’s explanation thus vindicates the intuition that physical depreciation is an insufficient reason for the sharp price differential between new and used cars. The gap is much more plausibly understood as a reflection of the fact that cars offered for



sale, taken as a group, are simply of lower average quality than cars not offered for sale.

### *The stigma of the newcomer*

The full disclosure principle also suggests why it might once have been more difficult than it is now to escape the effects of a bad reputation by moving. In the current environment, where mobility is high, a dishonest person would be attracted to the strategy of moving to a new location each time he got caught cheating. But in less mobile times, this strategy would have been much less effective, for when societies were more stable, trustworthy people had much more to gain by staying put and reaping the harvest of the good reputation they worked to develop. In the same sense that it is not in the interests of the owner of a good car to sell, it was not in the interests of an honest person to move. In generally stable environments, movers, like used cars, were suspect. Nowadays, however, there are so many *external* pressures to move that the mere fact of being a newcomer carries almost no such presumptions.

### *Choosing a relationship*

Most people want partners who are kind, caring, healthy, intelligent, physically attractive and so on. Information about physical attractiveness may be gathered at a glance. But many of the other traits people seek in a partner are difficult to observe, and people often rely on behavioral signals that reveal them. To be effective, such signals must be costly to fake. Someone who is looking for, say, a highly disciplined partner might thus do well to take special interest in people who run marathons in less than two and a half hours.

Even the degree of interest a person shows in a prospective partner will sometimes reveal a lot. Comedian and film star Groucho Marx once said he wouldn't join any club that would have him as a member. To follow a similar strategy in the search for a relationship would obviously result in frustration. And yet Groucho was

clearly onto something. There may be good reasons for avoiding a seemingly attractive searcher who is too eager. If this person is as attractive as he or she seems, why such eagerness? Such a posture will often suggest unfavorable values for traits that are difficult to observe. The properties of effective signals thus make it clear why coyness, within limits, is so adaptive. It is very difficult, apparently, for eager persons to disguise their eagerness.

The same properties also have implications for the institutional arrangements under which people search for partners. An often decried difficulty of modern urban life is that heavy work schedules make it hard for people to meet with another. In response, commercial dating services offer to match people with ostensibly similar interests and tastes. Participants in these services are thus spared the time and expense of getting to know people with whom they have few interests in common. They also avoid uncertainty about whether their prospective partner is interested in meeting someone. And yet while marriages do sometimes result from commercial dating services, the consensus appears, at least at present, to be that they are a bad investment. The apparent reason is that, without meaning to, they act as a screening device that identifies people who have trouble initiating their own relationships. To be sure, sometimes a participant's trouble is merely that he or she is too busy. But often it is the result of personality problems or other, more worrisome difficulties. People who participate in dating services are indeed easier to meet, just as the advertisements say. But signaling theory says that, on the average, they are less worth meeting!

#### *Conspicuous consumption as ability signaling*

Suppose one has been unjustly accused of a serious crime and is looking for an attorney to represent him. And suppose his choice is between two lawyers, who, as far as is known, are identical in all respects, except for their standard of consumption. One wears a threadbare polyester suit off the rack and arrives at the courthouse in a 15 year

old, rust eaten Chevy Citation. The other wears an impeccably tailored sharkskin suit and drives a new BMW 740i. Which one is more likely to get hired?

Signaling principles suggest that the latter attorney is probably a better bet. The reason is that a lawyer's ability level in a competitive market is likely to be mirrored closely by his income, which in turn will be positively correlated with his consumption. There is obviously no guarantee that the lawyer who spends more on consumption will have higher ability. But as in other situations involving risk, here too people must be guided by the laws of probability. And these laws say unequivocally to choose the better dressed lawyer.

Where important decisions involving people we do not know well are involved, even weak signals of ability are often decisive. Close employment decisions are an obvious example. First impressions count for a lot during job interviews. As the popular saying goes, "We never get a second chance to make a first impression." Placement counselors have always stressed the importance of quality attire and a good handshake in the job search process. Even when the employer *knows* how good an applicant is, she may still care a great deal about how that person will come across to others. This will be especially true in jobs that involve extensive contact with outsiders who do not know how good the employee is.

Judging from their spending behavior, many single people seem to believe that their marriage prospects hinge critically on what clothes they wear and what cars they drive. At first glance, this seems curious because by the time most people marry, they presumably know one another well enough for such things not to count for much. Even so, many potential mates have been rejected at the outset for seeming 'unsuitable'. The trappings of success do not guarantee that a person will marry well, but they do strengthen the chances of drawing a second glance.

The importance of consumption goods as signals of ability will be different for different occupations. Earnings and the abilities that count most among research professors are not strongly correlated, and most professors think nothing about continuing to drive a 15 year old automobile if it still serves them reliably. But it would be a big mistake for an aspiring investment banker to drive such a car in the presence of his potential clients.

This example makes it clear that a person's incentive to spend additional money on conspicuous consumption goods will be inversely related to the amount and reliability of independent information that other people have about his abilities. The more people know about someone, the less he can influence their assessments of him by rearranging his consumption patterns in favor of observable goods. This may help explain why consumption patterns in small towns, which have highly stable social networks, are so different from those in big cities. The wardrobe a professional person 'needs' in Iowa City, for example, costs less than half as much as the one that same person would need in Manhattan or Los Angeles. Similarly, because the reliability of information about a person increases with age, the share of income devoted to conspicuous consumption should decline over time. The more mature spending patterns of older people may say as much about the declining payoffs to ability signaling as about the increasing wisdom of age.

Note that conspicuous consumption as a signal confronts us with a dilemma. The concept of a tasteful wardrobe, like the notion of a fast car, is inescapably relative. To make a good first impression, it is not sufficient to wear clothes that are clean and mended. We must wear something that looks better than most others wear. This creates an incentive for everyone to save less and spend more on clothing. But when *everyone* spends more on clothing, relative appearance remains unchanged. In the familiar stadium metaphor, spectators leap to their feet to get a better view of an

exciting play. But if everyone leaps, the view is no better than if all had remained seated.

### **Signaling theory and its application to human signals**

Anthropologists have been pointing out the potential for animal signaling theory to form the basis of a rigorous, systematic, and scientific approach to human signals (e.g. Cronk, 2001; Harpending, Draper, & Rogers, 1987). That potential is now being realized thanks to several recent fieldwork-based studies of human signaling systems (see Bird, Smith, Alvard, & Chibnik, 2005 for a recent review; Cronk, 2005). However, as leading signaling theorists themselves have pointed out (e.g. Maynard Smith & Harper, 2003), terminological and theoretical confusion has slowed the development and use of signaling theory. Although signaling theory has much else to offer, there is no denying the centrality of costly signaling theory to this rapidly developing approach. Costly signaling theory seems likely to retain its importance both because it is a relatively well developed aspect of signaling theory and because it is useful for explaining signals that, due to their costs, are prominent, interesting, and attention-grabbing.

Anthropologists, following the lead of animal behavior studies, typically trace the idea that there is a relationship between the cost of a signal and its honesty to the work of Amotz Zahavi (1975; see also Zahavi & Zahavi, 1997) despite the fact that Zahavi's idea was anticipated several different times by social scientists. The list of social scientists and their ideas that are similar to Zahavi's includes Thorstein Veblen (1965) and the idea of conspicuous consumption, Thomas Schelling (1960) and his insights about signals of commitment, Michael Spence (1973) and his theory of job market signaling, and Robert Frank (1988) and his argument that moral commitments are hard-to-fake signs of one's reliability as a cooperator. Also, Bliege Bird et al. (2005) have pointed out similarities between costly signaling theory and

both Marcel Mauss's insights on competitive gift-giving (Mauss, 1954) and Pierre Bourdieu's (1977) idea of social capital.

The fact that costly signaling theory is common to both the social and biological sciences is more than just a curiosity. It also highlights the generality of signal design problems, whether they are solved by engineers, advertisers, or natural selection, and creates opportunities for fruitful exchanges of insights across disciplines.

Maynard Smith and Harper (2003) have recently made an attempt to clear up some of the terminological, conceptual, and theoretical confusion in animal signaling theory. Much of the confusion surrounds the relationship between honesty and cost. The emphasis on costly signaling may lead to the impression that costliness is a necessary guarantor of honesty. Although that is true in certain circumstances, it is not always so. For example, when signalers and receivers have common interests, selection (or signal design principles more broadly) may favor a signal that is only as costly as it needs to be in order to get the message across (Krebs & Dawkins, 1984; Maynard Smith & Harper, 2003). Such signals are said to have only "efficacy costs" (Guilford & Dawkins, 1991), i.e. the costs necessary to ensure that the information conveyed by the signal reaches the receiver. While efficacy costs can be substantial, they are not the sorts of costs referred to in the phrase "costly signaling theory". Costly signaling theory is concerned, rather, with strategic costs (Grafen, 1990a, 1990b), often referred to as handicaps. Strategic costs are necessary to ensure not that information is conveyed but rather that the signal is perceived as honest. The cosmetics study mentioned above provides an example of a signaling system with the potential for high efficacy costs but without strategic costs. Although women can spend a great deal of time and money on cosmetics, those costs are not related to the

qualities such as health, beauty, and youth that some women report that they are trying to convey with cosmetics.

Other signals may be honest not because they are particularly costly for the signaler but because they are simply impossible for those without the quality being signaled to successfully fake. Maynard Smith and Harper (2003; 1995) call these sorts of signals “indices”, which is related to the way that term is used in semiotics. For example, Hamilton and Zuk (1982) suggested that bare patches of skin on birds might serve as indices of their resistance to parasites, an idea supported by experiments on red jungle fowl (Zuk, Ligon, & Thornhill, 1992). Among humans, some markers of group membership, such as the ability to speak a specific dialect with the proper accent and complete fluency, have a similar quality of being either impossible or extraordinarily difficult to fake.

The literature on human signals also contains the category of “hard-to-fake” signals. This includes signals that are hard to fake either because they are indices or because they impose on signalers strategic costs which honest signalers can afford but which are difficult for dishonest signalers to bear. Sometimes the distinction between indices and costly signals is less important than the similarity of the circumstances in which signal design processes, including natural selection, may favor their development and maintenance. It is also sometimes helpful to avoid the word “costly” in order not to confuse strategic and efficacy costs. For both of these reasons, the “hard-to-fake” label is sometimes very useful.

One common feature of the literature on costly signaling is the idea that natural selection will favor costliness as a guarantor of signal honesty when there is a conflict of interests between signaler and receiver. A complementary idea is that when signaler and receiver have common interests, natural selection will favor expenditure on efficacy costs but none on strategic costs. In this section I argue that this

understanding of the circumstances that favor costly signals and, more broadly, hard-to-fake signals is not quite right. Specifically, costly signaling theory is relevant to circumstances in which there are broad conflicts of interests between categories of signalers and receivers but *confluences* of interest – common interests that are real though they may be fleeting – between *particular* signalers and *particular* receivers.

In evolutionary terms, a conflict of interests exists between two parties when natural selection would favor a different outcome for their interaction if it were determined solely by selection on genes in one party or the other (Maynard Smith, 1991; Trivers, 1974). The complement of this is that two parties share common interests when natural selection acting on genes in both of them would favor the same outcome from their interaction. Some categories of organisms are locked in permanent and perpetual conflicts of interests. Natural selection favors prey that can escape predators and predators that capture prey. It favors males that succeed in mating with many females regardless of their own quality and females that mate with only the highest-quality males. The relationships between other categories of organisms, such as parasites and hosts and parents and offspring (Trivers, 1974), are more complex but have potential for conflicts of interests. However, even in the context of such broad and permanent conflicts of interests, particular signalers and particular receivers can have common interests. It is in these situations that hard-to-fake signals will be favored. It is the difficulty of faking them, whether because they are indices or because they are costly, that ensures their honesty in a milieu in which honesty is not expected.

Stotting, a peculiar sort of hopping behavior performed by gazelles and some other ungulates when faced with a predator, is an example of a signal that is likely to have resulted from this evolutionary scenario. Field research has shown that the ability to stot correlates with an organism's physical condition and may help dissuade



predators from wasting time and effort in pursuit of an individual that is likely to escape capture (FitzGibbon & Fanshawe, 1988). While a prey species such as a gazelle and its main predators, such as African wild dogs, are certainly engaged in a long-term conflict of interests, a confluence of interests exists between an alert and physically fit gazelle that is capable of eluding a predator it spots and the predator. Natural selection would favor the same outcome from the interaction for both of them: abandonment of the pursuit. The gazelle saves the time and energy of eluding capture and the predator saves the time and energy of a failed pursuit. One might say that a conflict of interests still exists because natural selection would favor a successful hunt by the predators, but that is irrelevant. Only outcomes that are actually possible are relevant. Because the gazelle in question is alert and physically fit enough to avoid being caught, a successful hunt is so unlikely that it is not worth the predator's bother. At the beginning of the encounter, this information is possessed by the gazelle but not by the predator. Stotting is a way of transferring that information to the predator in a way designed to overcome the resistance to signals by the prey that selection has favored in it because of the broader conflict of interests between the two categories of organism. The conflict of interest in this scenario is not between an individual alert, physically fit gazelle and a specific pack of wild dogs but rather between an alert, fit gazelle and relatively inattentive, unfit gazelles in its vicinity.

A similar confluence of interests in the midst of a broader conflict of interests arises when a male that is of truly high quality relative to competing males attempts to convince a female to mate with him. Despite the broad conflict of interests between males in general and females in general, individual males and females can (and routinely do) experience confluences of interest. The male that is truly of high quality relative to competing males benefits from the encounter in an obvious way, i.e. by mating. The female benefits because she makes a good choice and mates with a male

that is truly of high quality compared to other available males. Hard to fake signals about male quality, whether they are indices or costly signals, serve to transfer honest information about the male's quality to a female that is likely to be skeptical due to past selection against females who made poor mate choices. As in the case of predators and prey, the conflict of interests in this situation is not between the female and the prospective male mate that is of truly high quality but rather between the high-quality male and low-quality males in its vicinity that would also like to mate with the female in question.

Such situations also arise routinely in human communication. For example, lobbyists use a variety of techniques in their efforts to influence people in government, some quiet and others elaborate (Baumgartner & Leech, 1998). Among the more dramatic types of lobbying is a grassroots campaign, in which an interest group encourages a large number of citizens to contact their legislators directly about a particular issue. Grassroots campaigns are one form of "outside lobbying", which contrasts with the "inside lobbying" style of personal contacts with legislators and their staffs (Kollman, 1998). Because "Astroturf" campaigns – fake grassroots campaigns mounted by interest groups that lack a large number of motivated members – are costly and difficult to organize, both policy makers and political scientists see grassroots campaigns as usually being honest indicators of how voters feel about issues (Kollman, 1998). Given the reasoning presented here, I would expect to find lobbyists using grassroots campaigns when trying to influence legislators with whom they are usually at odds. Such campaigns inform legislators that, although they may usually be opposed to the positions taken by the interest group, in this particular case there is a confluence of interests between the desires of the interest group to see certain bills passed or defeated and the legislators' desires to remain in office. This technique can be very effective. One congressional staff member explained why his

boss reversed his position on catastrophic health insurance in 1990 by explaining, “It was a no-brainer. He got over five thousand letters for the repeal of the insurance, and literally eight letters in favor of the current insurance. He didn’t have much choice really. He had to vote for repeal” (quoted by Kollman, 1998, p. 5).

The problems faced by job applicants signaling employers are analogous to those experienced by male organisms signaling potential mates. The broad context is adversarial, but if a particular job applicant is truly of high quality then there is a confluence of interests between him or her and potential employers. The details of job market signaling were explored by Spence (1973). Following the logic presented above, the most successful applicants will be those who honestly advertise their high quality with signals that would be too costly for low-quality applicants to fake. An example might be holding a degree with honors from an elite university (Frank, 1988, p. 102).

### **Signaling through digital artifacts**

Judith Donath talks about three types of signals in digital artifacts: 1) handicap signals, 2) index signals, and, 3) conventional signals (Donath, In Press). *Handicap signals* are costly to produce and are considered reliable because the quality they signal is ‘wasted’ in the production of the signal, and the signal tends to be more expensive to produce for an individual with less of the quality. An example of a handicap signal is active participation in online forums. An employee with over 10,000 forum posts proves that she has enough time to be active in the forum, while still maintaining her job responsibilities. She ‘wastes’ time to prove she has a surplus. “The Handicap Principle is a very simple idea: waste can make sense, because by wasting one proves conclusively that one has enough assets to waste and more” (Zahavi & Zahavi, 1997).

*Index signals* are directly related to the trait being advertised. These are reliable since they require that the sender possesses the relevant trait. For example, being a level 60 avatar, with accompanying powerful sets of armor and weapons in the popular multi-player online game World of Warcraft is an index signal. Another example is having a high number of positive ratings on the online auction site ebay<sup>8</sup>. Having the quality of being a good gamer is a pre-requisite to produce this signal. This connection between signal and quality makes an index signal reliable. Handicap and index signals are known together as *assessment signals*. Assessment signals relate to the quality it represents and thus one can assess the quality simply by observing the signal (Donath, In Press).

On the other hand, *conventional signals* are not correlated with a trait. The signaler need not possess the trait to send the signal. Because of this, conventional signals are less reliable and open to deception. The online world is rife with conventional signals. For example, it may be desirable to have an attractive picture of oneself on a social networking site such as MySpace. In the absence of social connections that can vouch for the veracity of such a picture, an individual may choose to put up a deceptive picture. If the use of such deceptive pictures becomes prevalent, the signal will lose its meaning as an indicator of attractiveness. Conventional signals are thus unstable because excessive deception can cause a once meaningful signal to turn into noise (Donath, 1999).

Signaling theory proposes that there are costs and benefits to both the sender of the signal and the receiver. For example, research has found that humans sometimes form automatic impressions on the basis of prior experiences (Greenwald & Banaji, 1995). However, when looking for expertise, one may want to engage in more detailed processing. Signaling theory provides an explanation regarding situations in

---

<sup>8</sup> <http://www.ebay.com>

which individuals may engage in automatic processing versus detailed processing. There is a concept of ‘receiver costs’ in signaling theory. If a reliable signal is very costly to assess, receivers might choose one that is less reliable but easier to obtain (Guilford & Dawkins, 1991). When the cost of making a poor decision is great, individuals will spend more time evaluating reliable signals and less time making automatic inferences. When the task at hand does not involve a high cost if a poor decision is made, individuals may engage in satisficing behaviors through automatic processing.

Given the proliferation of conventional signals online, it is not surprising that the majority of research on signaling in digital artifacts has been related to that type of signal. Donath looked at signaling in social networking sites such as Friendster and MySpace (Donath, 2007), where one might potentially artificially inflate his friends to appear popular or because of the social pressure to accept friend requests. Lampe, Ellison and Steinfield looked at another social networking site where users can selectively self-present themselves (Lampe et al., 2006). Investigating student behavior in the popular social networking site Facebook, they found that the completion of particular profile fields was a strong predictor of how many friends a student had. However, in the online world, assessment signals could be juxtaposed with conventional signals, albeit to a lesser degree. Inferred social connection information, as opposed to self reported social connection information which could potentially be deceptive, may act as an assessment signal of one’s sociability. In a similar way, expertise rank in an expertise locator system that is determined through an algorithmic process may act as an assessment signal of expertise. A contribution of this dissertation is to look at *both* assessment signals and conventional signals and how they are perceived.

## **A preliminary conceptual model of people sensemaking in expertise seeking behavior**

Signaling theory posits that there are costs and benefits to both the sender of the signal and the receiver. The basic equation of signaling theory states that a signal will be reliable when for honest signalers the benefits outweigh the costs while for deceptive signalers the costs outweigh the benefits (Donath, In Press). Based on the literature reviewed, I can construct a preliminary conceptual model of ‘people sensemaking’ in expertise seeking behavior. Because of factors prevalent in organizational settings such as increased accountability and non-anonymity, assessment signals may be more prevalent and deception may be low. Figure 5 shows a preliminary conceptual model of people sensemaking based on signaler costs and benefits and receiver costs and benefits. The signaler may use both assessment signals and conventional signals. The benefit to the signaler is that she may be able to influence the receiver’s beliefs through the signals she sends. The cost to the signaler is that the signal could be interpreted in a way that is exploitative or unintended. For example, within the context of expertise search, an individual might be sending an assessment signal of approachability through active participation in social software. But that signal may act as a double edged sword by inundating that individual with more expertise requests than she can handle. Eventually, depending on the goal of the individual, she might choose to tone down the assessment signal of approachability that she is sending.

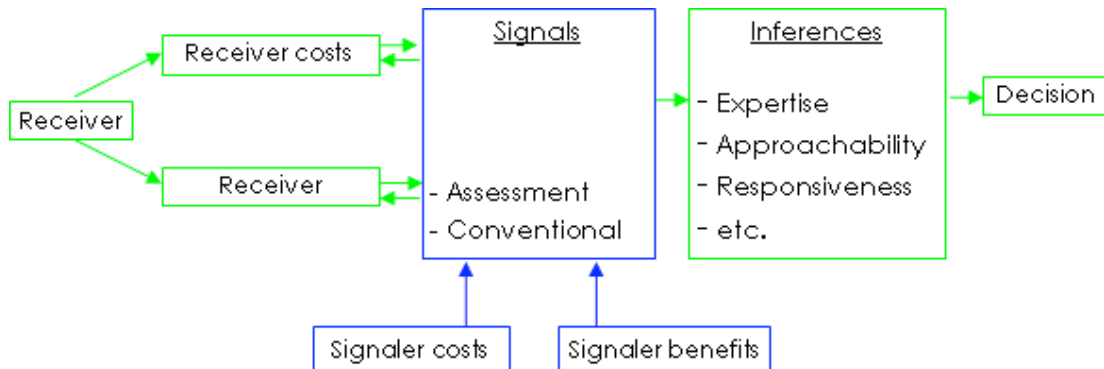


Figure 5. A preliminary conceptual model of people sensemaking in expertise seeking behavior. Items in blue correspond to the signaler and items in green correspond to the receiver.

The receiver of the signal also has benefits and costs. The benefit to the receiver is that they can comprehend the signal and modify their beliefs accordingly. Research has shown that individuals are particularly good at understanding and processing information about other humans when they first meet them (Uleman, 1999). This capability of making inferences may extend to the online world as well. Research in information retrieval has found that while performing an online search, some users do not quite know or are unable to articulate the object of their search. Yet they are able to recognize it immediately when they find it (Belkin, Oddy, & Brooks, 1982). Emerging research on how individuals comprehend online profiles seems to confirm this. Perceivers' personality trait ratings of Facebook profiles were strongly correlated with the users' self ratings and friends' ratings (Gosling et al., 2007). In addition, people believe that their Facebook profile represents them well (Lampe et al., 2006). There are findings that show that online profiles appear to represent individuals' offline personalities fairly well. Within an online dating context, Hancock, Toma and Ellison (2007) demonstrated that deception is minimal. But for a variety of reasons there may still be a large opening for manipulative signals. For example, human receivers may often lack sufficient information to accurately evaluate

the truthfulness of a signal. Religious admonitions to behave in particular ways in order to avoid an unpleasant afterlife are a dramatic case in point. Although one might find it hard to believe in supernatural beings and an afterlife, the possible cost of not believing in such things may seem so great that most receivers will choose to believe (or behave as if they believe, which to the signaler may be good enough). Similarly, the possible cost to a would-be rebel of assessing the truthfulness of a declaration by a political elite that certain types of behavior will be met by certain punishment in this life may be too great for it to be worth the risk. The ability of humans to use their signals to make claims that are difficult, risky, or even impossible to evaluate may help explain why animal signals are often so honest (Maynard Smith & Harper, 2003) while human signaling systems seem, to many observers, to be arenas for a great deal of deception and exaggeration (e.g. Alexander, 1975; Harpending et al., 1987).

The cost to the receiver in my model deals with the effort involved in evaluating a signal. If a reliable signal is very costly to assess, receivers might choose one that is less reliable but easier to obtain (Guilford & Dawkins, 1991). An aspect of animal signaling theory that has been underused by those studying humans is the importance of receiver psychology to signal design. Guilford and Stamp Dawkins (1991) argued that what receivers find easy to detect, discriminate, and remember constitute major forces in the evolution of signals. Noting that male birds frequently sing from high perches and that male frogs croak in frequencies suited to the ears of female frogs, they suggested that warning displays, such as the coloration of bees and wasps, might be designed not just to be conspicuous but also to be more easily remembered. Rowe (1999) expanded this idea to include multicomponent signals, noting that signalers can increase the chance that a message is received by sending it in more than one way.



The idea that signals are designed, whether by the signalers themselves or by evolutionary processes, in order to take advantage of receiver psychology is something that researchers in the human sciences have long understood. For example, mothers around the world speak to their babies in a special singsong, high-pitched fashion that researchers have labeled “motherese” (Fernald, 1992). Motherese differs from normal speech not only in terms of rhythm and intonation but also in that sentence structure is simplified and a lot of repetition is used. Interestingly, the babies themselves seem to prefer motherese to normal speech, paying more attention to people who speak in motherese than to those who speak as they would to adults.

Toy manufacturers are also keenly aware of the importance of appealing to the psychology of potential customers. This was nicely demonstrated by Hinde and Barden’s (1985) documentation of how teddy bears have evolved since they first became popular due to an association with Teddy Roosevelt. Teddy bears started out with prominent snouts, looking something like actual bears. Over the years, however, they shifted to a more baby-like appearance, with a reduced snout and a rounder face. Because the toy bears were being designed by their manufacturers to appeal to buyers, mainly adults, this evolution appears to reflect a process of signal evolution strongly influenced by a receiver psychology that includes a preference for babyish over more mature-looking faces. Perhaps that aspect of receiver psychology is particularly prone to activation when the receiver is shopping for a gift to give a child. Within the context of expertise search, in the self described expertise section of a corporate directory, an employee might list all the companies she has worked with while on assignment in Asia. However, if the receiver of that signal is not familiar with any of the companies listed, he will choose to focus his attention on other signals that might be easier for him to interpret.

In this dissertation, I will focus on the receiver of signals. Due to constraints of data collection, I was not able to interview any of the senders of signals (i.e. the profiles that study participants looked at). I will thus rely on empirical findings from recent research on individuals' behavior using social computing technologies to guide what motivations senders of signals in digital artifacts might have. While this is a limitation of the current dissertation, it is still important to understand how receivers perceive others, since that is the first step of the in the expertise seeking decision process. Research has also shown that perceptions of expertise are more influential than actual expertise in expertise seeking (Littlepage, Schmidt, Whisler, & Frost, 1995; Palazzolo, 2005).

## CHAPTER 3

### CONTEXT OF STUDY

#### **A look at expertise location/recommendation systems**

In looking at how people make sense of strangers' expertise using information retrieval systems, I will focus on software purpose-built for locating expertise, commonly known as expertise location/recommender systems. It is thus important to provide an overview of these systems. Terveen & McDonald provide a comprehensive review of what they refer to as 'social matching systems', which include expertise location and recommender systems (Terveen & McDonald, 2005). Some representative systems are briefly discussed below.

The Information Lens (Malone, Grant, Lai, Rao, & Rosenblitt, 1989) is a multi-user messaging system and represents an early attempt at solving the problem of information overload. The system contributed key ideas regarding cognitive filtering, social filtering and economic filtering techniques to identify messages which were particularly interesting to a user. The Information Lens is a type of electronic mail and bulletin board system shared among a group of participants. The system is organized around semi-structured text messages. Semi-structured text provides sets of standard key-value pairs that simplify parsing of the messages. Different types of messages have different sets of standard key-value pairs. For example, a memo might have structured fields for sender, date, title and space for a message body and an event message might have all the fields for a memo as well as additional fields for event location, event time, and event cost. Simple message types might have a small number of fields while more complex message types might have a large number of text fields.

Who Knows (Streeter & Lochbaum, 1988a, 1988b) is an early collaborative recommendation system that comes from the information retrieval tradition. Who

Knows attempts to solve the problem of finding an individual who can answer questions about a problem. This is distinctly different from Information Lens which focused on the content of messages. Who Knows recommends other individuals, not messages. The novel contributions of Who Knows are the application of Latent Semantic Indexing (LSI) techniques to both represent profiles and for matching queries. It also recommends people instead of messages or artifacts. Who Knows is unique among collaborative recommender systems in its use of LSI. In Who Knows, LSI is both the aggregation technique and part of the query technique. The concept of LSI as an aggregation technique might seem odd. But effectively the LSI profile of each individual can be thought of as a multi-dimensional score. In some dimensions a person is highly rated and in other dimensions they may have no rating at all. These multi-dimensional scores can be easily compared and ranked in order to make recommendations.

Tapestry (Goldberg, Nichols, Oki, & Terry, 1992), like Information Lens, is an attempt at solving problems of information overload. Tapestry and Information Lens support many of the same functions. Tapestry is a messaging system that supports email and bulletin boards through the use of semi-structured text messages. Posting, reading, replying to messages and filtering messages are all functions supported by Information Lens and Tapestry. However, Tapestry's capabilities surpass those provided by Information Lens in several ways. The messaging system in Tapestry includes two additional features. The system attempts to track when a user reads or views a message. The ability to track when a person reads or views a message provides an important event hook that can be used when filtering messages. Additionally, Tapestry provides an additional way to respond to messages. Tapestry supports a special purpose message called an annotation. Annotations are another type of semi-structured text message. Annotations can be viewed like full fledged

messages or restricted so that they are seen by only one or a small number of other people. Tapestry also supports a full fledged query language, Tapestry Query Language (TQL), with syntax and semantics much like SQL (Structured Query Language) for regular databases.

GroupLens (Konstan et al., 1997; Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994; Sarwar et al., 1998) follows in the footsteps of Information Lens with several improvements. The nature of the improvements have been to move the system from dependence upon dedicated Usenet news clients and GroupLens specific servers toward a more open system. The rationale for most of these improvements has been to allow GroupLens to engage more public recommendation environments. GroupLens differs from prior systems in several important ways. First, GroupLens is built around an open, previously existing messaging system, Usenet news. This differs from Who Knows which is a closed monolithic system. Secondly, GroupLens was designed as a highly distributed architecture. Tapestry also has a distributed architecture, but it was clearly designed for a local area network. The major difference between GroupLens and prior systems is the explicit collection of numeric evaluation or ratings of the messages read by a user. These evaluations are entered into a profile for the individual user. User profiles are compared looking for clusters of users who rate sets of articles similarly. Clustering is performed with a Pearson  $r$  correlation. These clusters are used to make predictive recommendations.

Active filtering (Maltz & Ehrlich, 1995) is a response to Tapestry, Information Lens, GroupLens and similar systems. Maltz and Ehrlich argue that prior systems are 'passive' in the generation of recommendations. Prior systems wait for a user to request a recommendation before providing one. Additionally, these passive systems recommendations are often lacking crucial context which makes the recommendation valuable. Maltz and Ehrlich argue that this is unnatural. They note that people tend to

hand each other recommendations, often without an explicit request. The active filtering system was constructed to solve some of the problems with passive filtering systems. In particular, active filtering was built to address ease of use and context problems through a direct, active, recommendation process. This recommendation system is not stand alone. Active filtering is an add-on to the Lotus Notes system.

PHOAKS (People Helping One Another Know Stuff) (Terveen, Hill, Amento, McDonald, & Creter, 1997) recommends URLs (Uniform Resource Locator) based on the number of unique times a URL is mentioned in a Usenet news group. PHOAKS specifically tracks individual URLs, but the intent of the system is to provide recommendations about the content that is represented by those web pages. PHOAKS uses a voting strategy as a way to aggregate evaluations. The system processes every Usenet news group message, looking for URLs in the text of each message. Each URL found in a message is considered a potential 'mention.' A URL mention is when the URL appears in the body of the message as new text, as opposed to in a reply or signature. Each unique mention of a URL is considered one vote for that URL. The primary means of getting recommendations from PHOAKS is through the PHOAKS website. The site is structured and organized using the same topic hierarchies as Usenet news. A user navigates the topic hierarchy looking for a desired topic. When the user finds the appropriate topic, she is given a list of the top 40 URLs mentioned in the equivalent Usenet news group.

Rooted in the field of CSCW, Ackerman and other researchers developed a series of systems that address both social and technical issues. Answer Garden (AG) (Ackerman, 1994) is a system designed to help in situations like technical support, where there is a continuing stream of questions, many of which occur repeatedly, but some of which have never been seen before. It has a branching network of diagnostic questions that helps users find the answers. If there is no available answer, it

automatically routes the question to the appropriate expert, then, the expert can answer the user as well as inserting it into the branching network. The design of AG addresses two important social issues in expertise finding. First, askers are anonymous to the experts, thus decreasing the asker's social psychological cost related to status implications and need for reciprocity. Second, by continually adding questions and answers into the corpus, it decreases the expert's workload in answering the same questions repeatedly as well as it grows an organizational memory incrementally. In the field study of AG, experts were manually selected, and there is not much direct port between askers and experts because of the anonymity. In a field study (Ackerman, 1998), Ackerman found these designs to be helpful. A number of users reported that it is beneficial to be able to ask questions anonymously. The other interesting finding is that a large proportion of the users did not get answers that were at the right level or length of explanation. This indicates that expertise systems should route organizational members more effectively to the right level of expertise instead of to the experts with the highest level of expertise. Furthermore, Pipek and Wulf (2003) applied the Answer Garden approach into different organizational setting. They found that incomplete data, continually changing classification schemes, and domain specific needs for technically mediation communications made adoption of an Answer Garden like system difficult. More importantly, they found that the Answer Garden approach is subject to the impact of the given division of labor and organizational micropolitics.

In Answer Garden 2 (AG2) (Ackerman & McDonald, 1996) an expertise location engine is provided. Various computer-mediated communication mechanisms are also added. AG2 also prefers to "stay local" when selecting expertise to allow contextualization and it supports an escalation process. Another interesting change of AG2 is that the system tends to blur the dichotomy between experts and seekers.

McDonald and Ackerman (1996) explained the reason as follows:

“While there was nothing in the underlying technology to force this dichotomy [in AG], it was a simplifying assumption in the field study to have separate user and expert groups. Real collectivities do not function this way. Most people range in their expertise among many different skills and fields of knowledge... We would like to allow everyone to contribute as they can, promoting both individual and collective learning.” (p. 98)

The Do-I-Care (DICA) agent (Ackerman, Starr, & Pazzani, 1997) was primarily designed to recommend ‘interesting’ changes in web pages to a single user. In this model, the user identifies a set of web pages that she wants to monitor and then trains a DICA to recognize changes to those web pages that the user thinks are interesting. DICA reads the contents of a web page and parses the page into chunks. A chunk is considered to be text-delimited by HTML tags. The set of chunks that compose the web page are then compared to a previously stored version of the same page. The differences between versions represent the changes that were made to the given web page. These changes are fed to a Bayesian classifier which determines which, if any, of the changes are interesting to the user. In a training mode, DICA can be trained to differentiate interesting and uninteresting changes.

ReferralWeb (Kautz et al., 1997a; Kautz, Selman, & Shah, 1997b) is another approach to finding a person that can answer a question. It analyzes public web documents to identify names associated with topics, uses co-authorship data to infer social relationships, and presents a referral chain showing the path from the seeker to the expert. ReferralWeb supports referral chains through the use of a social network. ReferralWeb models the social network incrementally based on the co-occurrence of names in publicly available web documents. The more frequently a set of people share references, the stronger is their social connection.



Yenta (Foner, 1997) is a matchmaking system designed to solve the problem of finding a person to answer a question. Yenta shares some similarity to Who Knows and ReferralWeb because they all attempt to solve the same problem. In many respects Yenta is a highly distributed version of Who Knows. However, Yenta does not rely on LSI techniques to analyze and cluster interests. Yenta is described as a personal information assistant. Each user who wants to get recommendations must have a Yenta. In the Yenta model each user sets up a profile of their knowledge by providing Yenta text based examples of their work such as email, reports and papers. This personal Yenta creates a profile of the local user by analyzing the text samples, creating a keyword vector for each sample and then clustering keyword vectors. The system makes recommendations through a network of communicating individual Yentas. When a user wants to find a knowledgeable person, he asks his personal Yenta by entering keywords as a query. The network of Yentas communicate to identify other people in the Yenta network who have a cluster of keyword vectors which most closely match the query. If a knowledgeable person is found, the Yenta network transports an anonymous question and anonymous response between the two parties.

Fab (Balabanovic, 1997; Balabanovic & Shoham, 1997) is a hybrid system that combines content and social filtering to make recommendations about web pages. Fab recommends web pages that a user will like based on the content of the page and other users who share similar interests. By combining content filtering and social filtering Fab claims to gain the power of both approaches while inheriting none of the common problems. The system partitions the recommendation problem into two steps. It first collects items into a manageable collection and then selects items from the collection to present as recommendations. It uses a collection stage to identify new or interesting web pages that should be added to the growing collection. Collection agents search

the web and select pages based on a cluster of similar users' profiles. Pages identified as interesting are recommended to users in the cluster.

Expertise Recommender (ER) mines software source control systems and technical support databases to associate specific individuals to specific software modules (McDonald & Ackerman, 2000). It then provides an instant messaging program to users logged into the system to contact individuals with knowledge of the modules. The system uses social networks to tailor recommendations to the user. The recommendation is done in two steps. First, it finds a set of individuals who are likely to have the necessary expertise. Then they are matched to the requester by using a social network. Expertise indication is defined according to organizational criteria. The technique for profile construction also depends on the organization. Experts are ranked according to expertise degree and social closeness. ER's main advantage is its flexibility towards expertise indications and sources. But identifying social networks in the organization can be very costly.

HALe (McArthur & Bruza, 2003) aims at discovering implicit and explicit connections between people by mining semantic associations from their email communications. Thus, email communication is the expertise indication in HALe. In this respect, it is similar to Yenta. Its approach to construct profile uses linguistic techniques. But HALe does not rank the experts and, therefore, the user has less information at hand to decide which expert to approach. Its transparency to the user is an advantage in HALe, since it does not disturb his/her activities. However, this system uses only one kind of information to indicate people expertise.

TABUMA (Text Analysis Based User Matching Algorithm) (Reichling, Schubert, & Wulf, 2005) generates users' profiles by extracting keywords from text documents. So, having text documents about a topic is the expertise indication in TABUMA. Notice that having the documents implicitly means that the person reads

them. Its approach to construct profile uses linguistic techniques. Experts are ranked according to their expertise degree. An advantage in TABUMA is its independence from text document format. A disadvantage is that it needs to ask users to choose a set of documents that they consider representative of their knowledge.

My research aims to add to this body of work by unpacking user behavior related to searching for expertise. While many of these systems attempt to identify an individual that possesses the expertise sought by a person, I believe there are other factors that need to be taken into consideration. For example, simply identifying an individual that has the knowledge a person seeks is fruitless unless that person actually responds. None of the systems reviewed in this section have an awareness of how approachable an expert is. There may be a few experts that could be consulted on a topic, but if they are not available then the expertise location process will not be successful. This dissertation will shed light on signals inherent in digital artifacts that convey a sense of approachability and responsiveness.

### **Study setting**

I conducted a field study at a global company specializing in information technology products and services. The company had various tools available to its employees that facilitated the search for people with expertise. These included 1) a corporate directory, 2) a blogging site, 3) a social tagging site, 4) an expertise locator, and 5) a dynamic directory that provided enhanced employee profiles (Farrell, Lau, Nusser et al., 2007). A user could type in a keyword into any of these systems and receive a list of people associated with that keyword. I conducted an informal investigation about the overlap of experts for a given expertise keyword search across these different systems. I typed in a sample expertise keyword 'AJAX' and analyzed the names of people associated with it. To measure overlap, I used the *binary overlap coefficient* (Manning & Schutze, 1999). The value of this co-efficient will be 1.0 when

all of the names are the same in both systems being compared. The results of my informal investigation are illustrated in Figure 6. As can be seen, there is very little overlap among the list of names returned by the five systems. The overlap in names between blogs and expertise locator was the highest (0.4). There was no overlap in names between directory and blogs, directory and tags, and blogs and dynamic directory. I take this as evidence that in a fairly large organization, there are a range of experts for a given topic (McDonald, 1999). Given this lack of overlap, I felt that my best option for obtaining a list of experts for a given topic was an expertise locator, since it was purposely built to find experts. I anticipated that it would provide me with a sample containing the majority of overlap in experts across different systems.

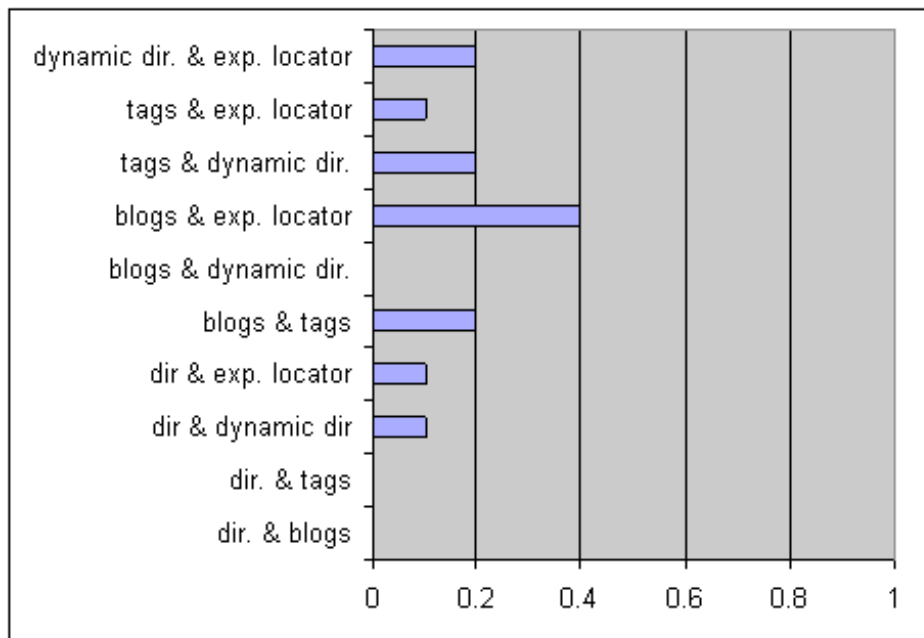


Figure 6. Overlap of AJAX experts across different systems

### **System used**

I used a recently developed prototype expertise locator system (Ehrlich et al., 2007; Lin et al., 2008) in this dissertation. The system analyzes the content of outgoing email messages and instant messaging transcripts to infer social connections and expertise. It runs a Google PageRank-like algorithm to associate names with topics to derive its expertise rankings (Lin et al., 2008). It allows users to search for individuals with specific expertise by typing in a query term. The system will display 10 names per page listed in rank order. The names are displayed in 5 rows, with 2 experts per row.

Figure 7 shows a sample results page for the search term “ajax”. For every person, there is a picture (A), name (B), business unit (C) and job description (D). One of the innovations in this system is that it also adds in the connection chain, up to 3 degrees, to indicate whether the person listed is a direct contact, 2 degrees away or 3 degrees away from the searcher. In Figure 7, (F) shows the person is 2 degrees away by displaying ‘Ask: [Person name]’, (E) shows that the person is 3 degrees away by displaying ‘Ask: [Person name] => [Person name]’, and (G) shows that the person is a direct contact by displaying ‘Your direct contact’. From this initial results list, users can click on any name to be taken to a page that contains more information about the person.

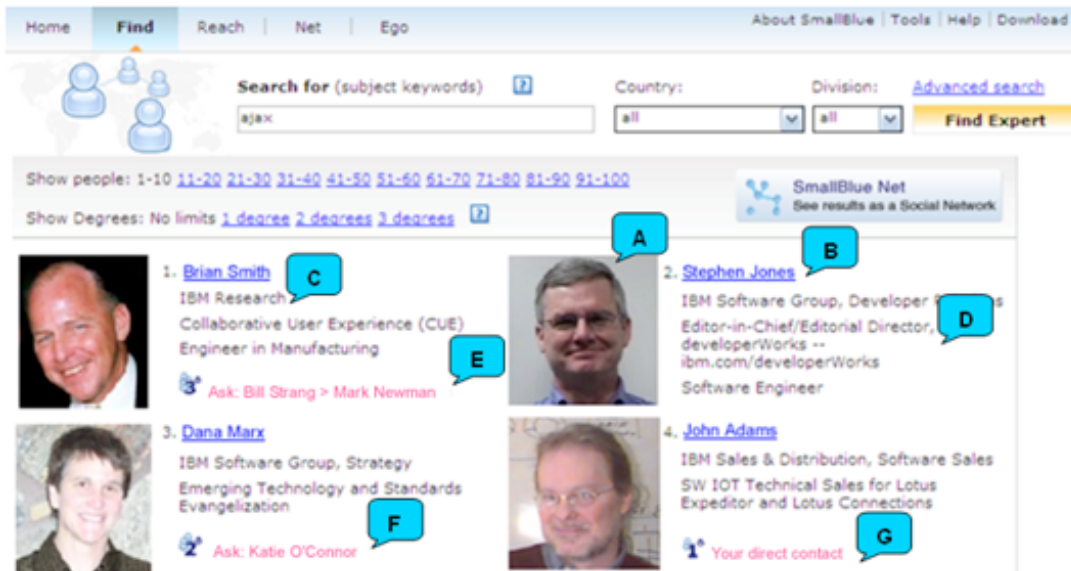


Figure 7. The first four of the top ten experts for AJAX. Pictures and names have been randomly used to protect privacy.

From this initial results list, users can click on any name to be taken to a dynamically generated profile of that person. A partial screenshot of a profile page is shown in Figure 8. The bottom left hand side of the page shows the top 30 social bookmarking tags followed by the number of times the tag has been used. On the right hand side of the page, the 5 most recent blog posts and their timestamp, the 5 most recent forum posts and their timestamp, and the 5 most recent bookmarks and their timestamp are displayed. The timestamps provide an indication of the recent activity level of the expert. In the middle of the page, the system displays a ‘recommended path’ from the user to the expert based on the shortest and strongest connection path calculated by the system. It also displays a list of alternate paths, up to six degrees away. If the expert is more than six degrees away from the user, the system will display a message stating ‘this person is more than six degrees away from you’. A more complete description of the algorithms driving the system and its user interface can be found in (Ehrlich et al., 2007; Lin et al., 2008).

James's community info      Your social paths to reach James      James's personal info

### Mailing list membership

Graph of people followed by James's

Informal Groups --

**In BlueGroups:**

- [Open Archives](#)
- [ResearchIT](#)
- [T&D Innovator](#)
- [Finance Software Expense](#)
- [TAMSESOPPhase2](#)
- [Social Networking Forum](#)
- [Reelive Beta](#)

[» show all Bluegroups James's in](#)

**In CommunityMap:**

- [Social Networking Forum](#)

### Social connection paths

Generate Paths

### Directory Info.

James H. Snell  
 M. Atom (Emerging Technologies, REST, XML, SOAP, and Web Services)

**Email:** [jhsnell@us.ibm.com](mailto:jhsnell@us.ibm.com)      Local time: 09:40 | 9:40AM

**Telephone:** 1-877-511-5082

**Notes:** James H Snell/Fresno/IBM@IBMUS

[Bluepages Profile](#)

### Social bookmarking tags

in Dogear (times James used it) (to tag a web-page, see the SN of who've used it)

- [innovation](#), 35
- [books](#), 11
- [lessons](#), 11
- [podcasts](#), 11
- [standards](#), 10
- [blogs](#), 9
- [extension](#), 9
- [unfiled](#), 9
- [html](#), 8
- [css](#), 8
- [web services](#), 8
- [standards](#), 7
- [coolness](#), 6
- [evolution](#), 6
- [cool](#), 5
- [opinate](#), 5
- [ubuntu](#), 5
- [atom](#), 4
- [java](#), 4
- [opensource](#), 4
- [research](#), 4
- [bank](#), 3
- [app](#), 3
- [hacks](#), 3
- [miscellaneous](#), 3

### Blog posts

What James talked about in public

**In Infrastructure Update** (29 Aug 2007)

- 1 [BCv2](#) (29 Aug 2007)
- 1 [BCv2 Poll - Please read](#) (08 Aug 2007)
- 1 [The Atom Protocol is Done](#) (28 Aug 2007)
- 1 [Infrastructure Update](#) (29 Aug 2007)

### Forum posts

**In IBM forums:**

- 1 [What happened to Date under advanced Settings when posting a blog entry?](#) (19 Aug 2007)
- 1 [Can I be longer able to use "I" in Blog Entry Titles?](#) (19 Aug 2007)
- 1 [Can I embed Flash in a blog entry?](#) (19 Aug 2007)

### Social bookmarks

**In Dogear (bookmarks):**

- 1 [Social bookmarking callation](#) (23 Aug 2007)
- 1 [Bookmarks](#) (19 Aug 2007)
- 1 [xml:object](#) (19 Aug 2007)
- 1 [mnotes' s Web log: US! Templates Redux](#) (19 Jul 2007)

### Self described expertise in corporate directory

James's descriptions in Bluepages

The following are emerging technologies, REST, XML, SOAP, and Web Services

standards, Best practices, Technical communication, Social Networking, Wiki's, Podcasting, Emerging Technologies, Web Services

**Technology Experiences:**

- Commerce
- Enterprise
- Finance
- Interfaces
- User Interfaces
- Wireless Networking
- Procedural Languages
- IDEs (Integrated Development Environments) & GUI
- Open Source Software Movement
- Computer Science
- Object-Oriented Languages
- HTTP Servers
- Autonomic Computing
- MIME (Multipurpose Internet Mail Extensions)
- Database Software & Middleware
- Markup Languages
- Web Application Servers
- Web Application Servers
- Web Application Servers
- Interpreted Languages
- Web Application Servers
- Communications & Networking Software
- Computer Security & Encryption Software

Figure 8. Screenshot of a 'profile' page. Pictures have been obscured to protect privacy.

It is noteworthy that this study is not an evaluation of the expertise locator system. I used this system because it provided a convenient research platform that allowed me to look at various signals of interest (expertise, social closeness, geographic distance, participation in social software) aggregated together in a single place.



## CHAPTER 4

### EMPIRICAL STUDY OF EXPERTISE SEARCH

#### **Making sense of initial expertise search results pages**

At least superficially, searching for experts is similar to searching for web pages. In both cases, the search results usually contain a link to a personal web page or email address, accompanied by the name, a picture, and a “snippet” of summary information about the person or web page. Previous research on web searches has highlighted the importance of accompanying information such as captions (Clarke, Agichtein, Dumais, & White, 2007), snippet length (Cuttrell & Guan, 2007) and rank order of results (Guan & Cutrell, 2007) for which items users will select for further exploration. Given the growth of expertise location tools, it is worth exploring user behavior within the context of expertise search. In this chapter, I will describe two studies in which I examined the factors that predict the likelihood of clicking on a particular search result within a results page (Master page) for further exploration. There are certain signals embedded within a results page that may influence clicking behavior.

#### **Hypotheses**

Much of the work on user selection from search results has looked at searches for documents. Clarke et al. (Clarke et al., 2007) looked at the influence of captions - the title, URL, and snippet of text that summarized the contents of the page. They analyzed logs of the Windows Live search engine and found that relatively simple features such as presence of query terms, readability of the snippet and length of the URL significantly influenced clickthrough patterns. In an eye-tracking study, Cuttrel & Guan manipulated snippet length for informational and navigational searches and found that longer snippets led to an increase in performance for informational searches but a decrease for navigational searches (Cuttrell & Guan, 2007). In the same eye-

tracking study, Guan & Cuttrel looked at the effect of rank order on informational and navigational searches (Guan & Cuttrel, 2007). For both types of searches, they found that there was a decrease in click rates as most users only focused on the first few results at the top of the page. Such findings are similar to eye tracking studies which revealed a bias for higher ranked results, even when the snippets of those results were less relevant (Joachims et al., 2005; Pan et al., 2007).

To the best of my knowledge, there is no comparable research on searches for people. However, it is reasonable to think that rank order matters since relevance is closely related to skill and other matches in expertise search engines (Ehrlich, 2003; Terveen & McDonald, 2005). With respect to people search, Fiore and Donath found that users of an online dating site preferred similar others when looking for a romantic partner (Fiore & Donath, 2005). However, they could not look at the intermediate step of selecting potential dating partners from the set of search results displayed. The literature on social ties is mixed when it comes to predicting the effect of social connectedness on link selection. Some research suggests that people might go to weak ties because such ties provide information different than that found in one's own social circle (Granovetter, 1973). Yet other studies show that people go to others they know directly (Nardi, Whittaker, & Schwarz, 2002) and that weak ties are adequate when seeking technical information (Constant et al., 1996). Still other studies have suggested that people will go to those they know well for complex information but weak ties are sufficient when the information is not complex (Hansen, 1999).

Both expertise rank and social connection information could be thought of as assessment signals. As opposed to self reported conventional signals, both these pieces of information are inferred by SmallBlue. This leaves little opportunity for deception. I thus hypothesize that expertise rank and social connection information will predict the likelihood of clicking on a particular link for further exploration.

Hypothesis 1: *Higher expertise rank will be positively related to clicking on a search result for further exploration.*

Hypothesis 2: *Existence of social connection information will be positively related to clicking on a search result for further exploration.*

### **Phase 1, Part A: User study with a single keyword**

#### ***Participants***

Sixty seven full time employees located in 21 different countries that had performed at least 20 searches using SmallBlue participated in my study. Majority of participants were from the United States and worked in the services division of the company. Their average tenure was 10.5 years.

#### ***Procedure***

Due to the geographic spread of participants and ease of setup, I conducted this study over the phone. Conversations were recorded with the permission of participants.

Each participant was instructed to imagine they were on a committee evaluating a new project proposal that was proposing to use AJAX for part of the project. They had to find an expert who could provide a second opinion on the suitability of using AJAX. I chose AJAX as the query term since it was one of the most frequently searched keywords found in the search logs of SmallBlue. As the participant entered the search term, the researcher would do the same. Anyone typing in the same search term in SmallBlue will see the same results. Only the social connection information is personalized to each user.

Once the results appeared, participants were given time to review the set of names. The researcher then asked which of the 10 experts, displayed on the first page, the participant would like to find more information about. There was no limit on the number of choices.

Unlike prior studies, I did not use a proxy to manufacture search results. In keeping with the spirit of a field study, participants were also not instructed to stay within the list of top ten experts. Participants that either a) searched beyond the initial list of 10 experts, b) used one of the available filters to search within 1, 2 or 3 degrees, c) used keywords in addition to AJAX e.g. AJAX user interface resulting in a list of people different than the ones displayed in Figure 7, and d) used a filter to search within a particular geography were excluded from the study in order to facilitate data analysis on a consistent set of 10 experts. These 10 experts were the same throughout the study.

### ***Measures***

#### *Whether a person is considered*

My dependent measure was a dichotomous variable measuring whether or not a participant clicked on a name to get more information on that person. I did not have access to live log data and relied on the participant telling us who they selected.

#### *Rank order*

The rank order of experts in the search results were coded as a categorical variable with 5 levels representing the 5 rows of the search results. This variable was then dummy coded with row 5 as the base category.

#### *Social connection information*

I coded social connection information categorically as either present, if there was a connection of any degree, or absent. Forty (59.7%) of my participants had social connection information displayed for at least one expert. Nine (13.43%) knew at least one expert directly. There was no correlation between rank order of expert and having social connection information displayed ( $r = 0.061$ ,  $p = 0.12$ ).

### *Familiarity with AJAX*

Upon completion of the study, participants were asked to rate their familiarity with AJAX on a scale of 1 to 5 where 1 = I have not heard of AJAX before, and 5 = I use it regularly. This was used as a control variable. The average rating was 3.81 with the majority of participants reporting that they had heard of AJAX but had no training in it.

### **Results**

Each participant had 10 choices to consider. A choice of the same participant could be related to her other choices. Choices were thus clustered by participant, making the observations non-independent of each other. To account for this, I analyzed data using the generalized estimating equations (GEE) method (Norton, Bieler, Ennett, & Zarkin, 1996). GEE controls for within-cluster correlation in regression models with binary outcomes. The results of the analysis are summarized in Table 2. The odds of considering a person increase roughly 5.5 times when going from row 5 to row 1 in the result set ( $\beta = 5.64, p < 0.001$ ). Hypothesis 1 was thus supported. The odds of considering a person increase roughly 4 times when there is social information available in the snippet ( $\beta = 3.93, p < 0.001$ ). Hypothesis 2 was thus supported. Familiarity with AJAX ( $\beta = 0.2, p = 0.27$ ) dropped out of the model. Figure 9 shows the number of times an expert was considered as a function of their rank order in the search result list.

Table 2. Results of GEE for Phase 1, Part A. Only significant predictors shown. Note: N = 670, \*\*p < 0.001

	B (SE)	95% CI for Exp (B)		
		Lower	Exp (B)	Upper
Constant	- 2.5** (0.41)			
Social Info.	1.37** (0.26)	2.4	3.93	6.5
Row 1	1.73** (0.42)	2.46	5.64	13.1

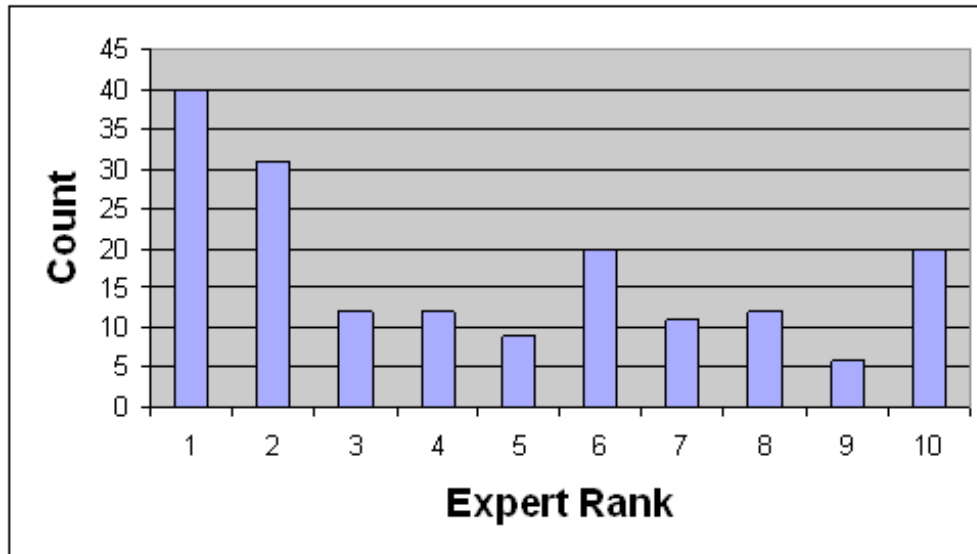


Figure 9. Number of times an expert was considered in Phase 1, Part A

The significance of rank order and social connection information signals might be because majority of participants were not familiar with AJAX and thus put more trust in the system. To determine whether my findings hold across different expertise keyword searches and consequent list of different experts, I conducted part B of phase 1 by varying the expertise keyword searched, resulting in a ‘random’ list of experts that participants saw.

### **Phase 1, Part B: User study with multiple keywords**

#### ***Participants***

Seventy five full time employees located in 21 different countries that had performed at least 20 searches using SmallBlue participated in my study. This was a larger pool than part A since all participants followed instructions and no participant was excluded. Majority of participants were from the United States and worked in the services division of the company. Their average tenure was 10.5 years.

#### ***Procedure***

Similar to part A, this study was conducted over the phone and conversations were recorded with the permission of participants.

The use of scenarios is a widely adopted method for investigating how individuals interact with technology (Carroll & Rosson, 1992). Terveen & McDonald suggest using scenarios that are specific to the participants’ tasks and organizational settings (Terveen & McDonald, 2005). Following their recommendations, I had my participants imagine themselves in the following scenario.

*“I want you to reflect back on a situation during your career at [company name] where you needed to locate people that have expertise on a certain topic. I’ll give you some time to think about this expertise. Once you’ve thought about it, let me know the expertise keywords you would use to search for a person with that expertise.”*

The researcher would then ask the participant to provide a rating on a scale of 1 to 9 (where 1 = not important at all and 9 = extremely important) regarding how important it was for the participant to find the right person to contact. The mean rating on this scale was 8.08 ( $SD = 1.47$ ), indicating the high importance of finding the right expert.

Participants were then told to enter the keyword they would use to search for a person with the expertise they sought. Expertise keywords entered by participants had a wide variety, but were mostly related to technology. Similar to part A, as the participant entered the search term, the researcher would do the same. Anyone typing in the same search term in SmallBlue will see the same results. Only the social connection information is personalized to each user.

Once the results appeared, participants were given time to review the set of names. The researcher then asked which of the 10 experts, displayed on the first page, the participant would like to find more information about. There was again no limit on the number of choices.

### ***Measures***

#### *Whether a person is considered*

My dependent measure was a dichotomous variable measuring whether or not a participant clicked on a name to get more information on that person. I did not have access to live log data and relied on the participant telling us who they selected.

#### *Rank order*

The furthest my participants went in exploring the result set was the sixth page. This created a list of 60 experts. With 2 experts per row, these experts were coded into 30 tiers. Majority of participants considered the first two pages, which displayed experts from tiers 5 to 10. Since the remaining tiers did not have many data points, I entered this variable as a continuous variable in my model.



### *Social connection information*

I coded social connection information categorically as either present, if there was a connection of any degree, or absent. Fifty eight (77.33%) of my participants had social connection information displayed for at least one expert. Thirty seven (49.33%) knew at least one expert directly.

### **Results**

Similar to part A, a choice of the same participant could be related to her other choices. Choices were thus clustered by participant, making my observations non-independent of each other. To account for this, I analyzed data using the generalized estimating equations (GEE) method (Norton et al., 1996). GEE controls for within-cluster correlation in regression models with binary outcomes. The results of the analysis are summarized in Table 3. The odds of considering a person increase slightly above one times when going from the last result page to the first result page ( $\beta = 1.1$ ,  $p < 0.001$ ). Hypothesis 1 was thus supported. The odds of considering a person increase roughly 3.5 times when there is social information available in the snippet ( $\beta = 3.39$ ,  $p < 0.001$ ). Hypothesis 2 was thus supported. Figure 10 illustrates how many times different experts were considered across rank tiers.

Table 3.  
Results of GEE for Phase 1, Part B. Note: N = 1070, \*\*p < 0.001

	B (SE)	95% CI for Exp (B)		
		Lower	Exp (B)	Upper
Constant	0.22 (0.14)			
Social Info.	1.22*** (0.2)	0.82	3.39	1.62
Rank order	0.08*** (0.02)	0.03	1.1	0.13

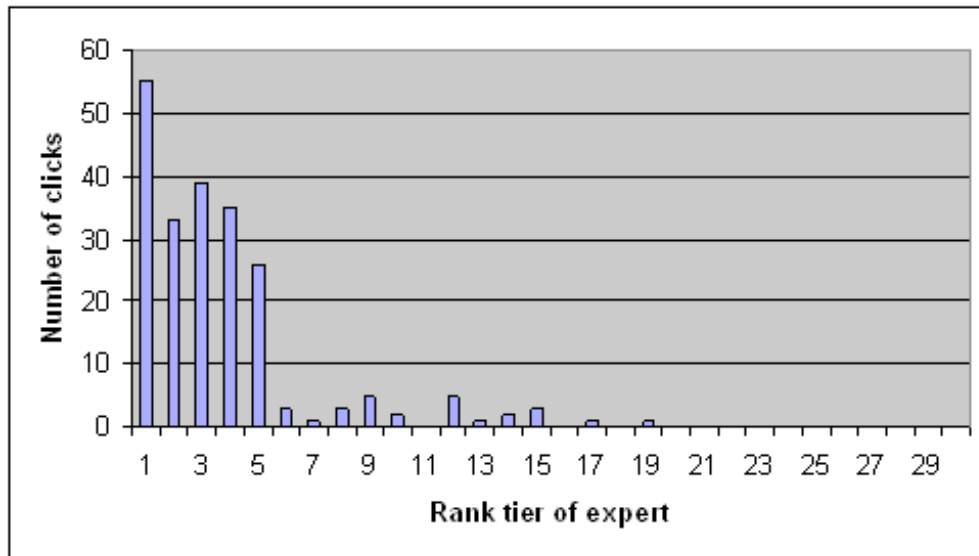


Figure 10. Number of times an expert was considered in Phase 1, Part B

### ***Discussion***

My results indicate that when considering experts, people prefer others they share a social connection with over a complete stranger. This has important implications for expertise search. Prior research suggests that interacting with those outside one’s social circle provides access to different and unique perspectives (Granovetter, 1973). However, my participants did not consider experts that were more than 3 degrees away who could potentially be a source of diverse expertise. Social context outweighed the potential of obtaining diverse expertise among my participants. Although prior studies have suggested the benefits of using social information (Terveen & McDonald, 2005), ours is the first to empirically demonstrate the role of social connections in selection choices.

Interestingly in part A, majority of participants did not select names of people they knew directly (of which there were very few), since the profile page would not have provided any additional information. Information regarding who were 2 or 3 degrees away was thus very influential in link selection decisions. This has important

implications for the design of displaying results in expertise locator systems. Individuals valued information about who they have a connection path to. A system that makes this information explicit and easily available is thereby increasing its utility in the eyes of users. When looking for specific expertise, if ‘name dropping’ of mutual acquaintances increases common ground and the probability of response from an expert, then displaying information regarding which expert one has mutual acquaintances with is extremely valuable.

The results also show that rank order predicted whether a search result was considered for further exploration. This was obtained in both part A and part B. This is consistent with prior research on document search which shows a bias towards selecting search results higher in the list (Guan & Cutrell, 2007; Joachims et al., 2005; Pan et al., 2007). Thus this study extends prior studies of web searches to show that some of the same effects, namely rank order, hold when looking for people. However, other factors, in this case, social connection, indicate there may be additional factors to consider in expertise searches. Searching for experts may superficially look like any other kind of search but searching for people takes place in a social context such that the relationship between the searcher and the expert is an important signal that influences the decision process (Ehrlich et al., 2007; Terveen & McDonald, 2005). Taking into consideration the fact that rank order did not exert much influence in part B simply amplifies the significance of social connection information as a signal that predicts likelihood of clicking behavior.

### **Making sense of profile pages of experts**

In previous chapters I have discussed how signals are prevalent in digital artifacts and which signals are important in an initial search result set. However, the signals in a search result set are somewhat limited. A profile that aggregates information from different sources may have many more signals that one can interpret.

Gosling et al. (2002) explain two mechanisms through which people form impressions of other's in a physical environment: a) identity claims and b) behavioral residue. Identity claims draw on the notion of impression management (Goffman, 1959) with regard to how a person would like to be perceived. Identity claims have the potential of self-presentation since these claims can be manipulated. On the other hand, behavioral residue refers to traces of behavior left behind without explicit intentions of self presentation. The challenge with identity claims and behavioral residue is that they are confounded. It is hard to separate deliberate self presentation from inadvertent self expression. Vazire & Gosling (2004) extend Gosling et al.'s (2002) model to the online world and show how identity claims and behavioral residue are also present in the digital realm. Specifically, they looked at how personal webpages increase the opportunity for identity claims while minimizing behavioral residue.

In this phase of the study, I have a unique opportunity to look at both self-authored and other-authored content. Inferred social connection information, as opposed to self reported social connection information which could potentially be deceptive, may act as an assessment signal of one's sociability. In a similar way, expertise rank in an expertise locator system that is determined through an algorithmic process may act as an assessment signal of expertise. Finally, participation in various forms of social software may provide signals of approachability and responsiveness. In this phase, I look at the juxtaposition of assessment signals and conventional signals, how they may influence whom a person decides to contact and how their meanings are interpreted.

## **Signals influencing whom a person decides to contact for expertise**

### ***Social software as a signal of approachability***

One of the main motivations of this dissertation was to investigate the influence of participation in social software (e.g. blogs, social bookmarking and tagging, online forums) and its relation to expertise search. Many online profiles incorporate some form of social software such as links to a blog or display of social connection information. Recent research on social software has shown they can be used for altruistic and community development purposes, such as help in obtaining expertise. In a survey of social tagging systems, Marlow et al. found that one of the motivations behind tagging is “contribution and sharing”, defined as tagging for either known or unknown audiences (Marlow, Naaman, Boyd, & Davis, 2006). Many of the popular blogs on the internet are frequented because people are interested in the opinions and expertise of those bloggers. Sites such as Yahoo! Answers and similar online forums provide a platform where people can access the expertise of others. While some of these sites may or may not be anonymous, recent work in organizational settings has demonstrated how social software such as social bookmarks and tags, and dynamic directories (Farrell, Lau, Nusser et al., 2007) can help locate non-anonymous experts. Millen et al. describe how their Dogear enterprise social bookmarking service can be used to locate experts through direct access to a person’s email, personal page, and blog from their bookmarks (Millen et al., 2006). While users can choose to make their bookmarks private, making bookmarks public may suggest a desire for self-presentation (Goffman, 1959). Thom-Santelli et al. found that indeed social tagging behavior was motivated by awareness of an audience and the need to communicate with them to build community (Thom-Santelli, Muller, & Millen, 2008). Participation in social software may also send out a signal of approachability. Organization members face a choice between sharing their

knowledge and interests with others and keeping it private. Those willing to share may be signaling their approachability to others. Farrell et al. found that an enhanced corporate directory was used to create community and introduce people to each other through people tagging (Farrell, Lau, & Nusser, 2007). Users frequently tagged people for the benefit of others. They wanted users to find out about each other and encourage them to start using the tagging feature. Since reciprocating requests is a hallmark of a community, participation in such people tagging could be conceived as identification of people willing to respond to expertise requests. While research in this area is clearly still emerging, I may be able to hypothesize that users of social software have an audience in mind when they participate in these systems, and they may be sending out a signal to others that they are approachable for contact.

Signaling theory could be used to explain this seemingly ‘irrational’ behavior of participation in social software within a corporate setting. In their 2005 article, Bliege-Bird and Smith cite Thorstein Veblen (1965) and Marcel Mauss (1954) as among the first to identify the underlying motivations for ‘irrational’ economic activities. These explanations focus on the elevation of prestige, status, and reputation through costly communication. For example, Thorstein Veblen’s (1965) theories of ‘conspicuous consumption’ and ‘conspicuous leisure’ identify how information about wealth is communicated, resulting in higher prestige. The acquisition of wealth and power is not sufficient, as these gains must be advertised to others to gain esteem (Veblen, 1965). Because wealth is commonly acquired through labor, an individual who conspicuously does not labor, but instead pursues ‘quasi-scholarly or quasi-artistic’ endeavors, demonstrates that they need not labor for wealth. In fact, when there is ‘industrial differentiation of classes,’ labor is viewed as vulgar and becomes taboo. Conspicuous displays of wealth through dress, eating habits, avoidance of labor, and occupation of large homes with expansive servant quarters, are indicative of

‘pecuniary strength’ (Veblen, 1965). Veblen writes that “unproductive consumption of goods is honourable, primarily as a mark of prowess and a prerequisite of human dignity” (p. 69), and hypothesized that individuals with ‘new wealth’ are more likely to participate in such conspicuous displays of irrational economic behaviors because there is a degree of uncertainty involved in verifying wealth. Individuals from known wealthy families may be less likely to participate in such displays because of common knowledge of financial standing and little need to demonstrate their wealth (Bird et al., 2005; Veblen, 1965). The ability to demonstrate that financial gain or that earning a living is unimportant indicates that the consumer does indeed possess significant wealth. This should increase prestige (Veblen, 1965).

Other early works, such as Marcel Mauss’ *The Gift* (Mauss, 1954), analyze aspects of conspicuously ‘wasteful’<sup>9</sup> economic behaviors. According to Mauss’ interpretation of gift giving, particularly in ritualized settings such as the potlatch among the Northwest Coast Indians, conspicuous consumption and conspicuous gift giving may have similar outcomes. In both cases, the ability to dispose of resources ‘recklessly,’ may serve as a means of enhancing the prestige of the giver. Mauss (1954) rejected previous notions of profit maximization through the potlatch, instead asserting that the wealthy man could build prestige and honor through virtually unlimited giving and destruction of wealth, demonstrating his ability to destroy valuable resources (Bird et al., 2005; Mauss, 1954).

Similar to signaling through handicapping physical features such as colorful plumage or ornamental antlers in the animal world or rejecting the means of acquiring wealth through work in humans, there may exist signals in digital artifacts that are difficult to fake, and allow inferences of expertise. For example, participation in various forms of social software such as blogs and online forums could be considered

---

<sup>9</sup> Here ‘wasteful’ is not used in a pejorative sense, but in the ‘less economically rational’ sense

a 'wasteful' activity. In organizational settings, employees are not paid to blog or participate in forums, and the opportunity cost of such participation leaves employees less time to focus on their primary task. However, employees could be participating in social software with the aim of creating social capital through such seemingly irrational and 'wasteful' activities. These considerations suggest the following hypothesis.

Hypothesis 3: *Participation in social software will be positively related to likelihood of considering contacting someone.*

### ***Social closeness***

Social closeness is conceptualized in terms of the social network concept of tie strength (Granovetter, 1973). Typically, a person develops ties to others she spends time with, and shares emotional intensity, intimacy and reciprocal services. She is socially distant to others she would have to go through many intermediaries to connect to. While weak social connections provide 'information benefits' through access to novel information (Granovetter, 1973), social closeness enhances cooperation and open communication (Jehn & Shah, 1997) possibly because of the emotional attachment (Brass, 1992), and intimacy (Wiseman, 1986) that are intrinsic to the relation.

Research has found that social closeness tends to develop between people who share commonalities, including race and gender (Ibarra, 1992). Such commonalities or homophily helps individuals understand each other better. People that are socially close tend to have developed a relationship-specific heuristic for processing tacit knowledge between them. Because of shared understanding, perceived social similarity and frequent interaction, socially close individuals may not need to spend much effort sharing tacit knowledge. In addition, social closeness often allows multiple interactions between people (Leonard-Barton & Sinha, 1993). Individuals



have the opportunity to try, err, and seek instruction and feedback. In contrast, among socially distant individuals, the necessary interactions for transferring complex knowledge will require a lot more time and may even become burdensome. Based on the above, I can arrive at the following hypothesis.

Hypothesis 4: *Social closeness will be positively related to the likelihood of considering contacting someone.*

### ***Quality of expertise***

Cross and Borgatti (2004) found that quality of expertise “is perhaps the single most important variable in knowledge seeking” (p. 153). Research has shown that the higher the quality of perceived expertise of a person, the more likely individuals will contact that person for expertise (Morrison & Vancouver, 2000). Palazzolo found that organization members are highly likely to retrieve information from those whom they perceive as experts for a given topic (Palazzolo, 2005). This is not to say relational factors such as accessibility do not matter, but results from prior studies reveal the strong tendency of people to obtain expertise from others they perceived to be knowledgeable in related knowledge areas. Since expertise location tools typically display search results in a ranked order, research on the effect of rank order on search result selection patterns becomes relevant. Within document search, numerous studies have shown the strong effect of rank order on link selection (Cutrell & Guan, 2007; Granka et al., 2004; Joachims et al., 2007; Pan et al., 2007). To the best of my knowledge, only a single study has looked at the effect of rank order within expertise search. That study found that the higher the rank order of an expert, the higher the likelihood a person will select that expert (Shami, Ehrlich, & Millen, 2008). Based on these findings, I propose the following hypothesis with regards to contacting a person for expertise.

Hypothesis 5: *Expertise rank order is positively related to the likelihood of considering contacting someone.*

### **Phase 2, Part A: User study of profile pages of experts on a single topic**

Phase 2 of the study involves two user scenarios. In part A, participants are asked to identify an expert in AJAX to contact. Part A was designed to answer the question: when looking at a profile page of an expert, what factors predict the perceived likelihood of contacting that person. In part B, participants are asked to reflect back during their career when they needed to contact someone for specific expertise. They were then asked to use SmallBlue to find an expert on that topic. Part B investigated the individual contribution of different pieces of information within the context of seeking to contact someone for specific expertise. The order that participants participated in part A and part B of phase 2 was counter-balanced to negate any effects of one scenario influencing the other.

### ***Participants***

Sixty seven full time employees located in 21 different countries that had performed at least 20 searches using a prototype expertise locator system participated in my study. Majority of participants were from the United States (43.75%), followed jointly by the United Kingdom (11.25%) and Canada (11.25%). There were 48 males and 19 females. A majority of participants (37.5%) were from the business services division of the company. Their average tenure at the company was 10.5 years. Of the participants, majority (33.33%) reported using the system at least once a month. Participation in my study was not contingent on frequent use of the system. I was interested in individuals that had a declared need for searching for people, as demonstrated through voluntarily performing over 20 searches in the system.

## ***Procedure***

The use of scenarios is a widely adopted method for investigating how individuals interact with technology (Carroll & Rosson, 1992). Terveen & McDonald (Terveen & McDonald, 2005) suggest using scenarios that are specific to the participants' tasks and organizational settings. Following their recommendations, I had my participants imagine themselves in the following scenario and asked them to try to act as if they are experiencing it in real life.

*“You are on a committee that is evaluating a new project proposal. One of the other committee members has remarked that the proposal is making inappropriate use of AJAX to implement a portion of the user interface. AJAX is a web development technique that enables many of the Web 2.0 style interactions. You don't know AJAX yourself but you decide to seek an AJAX expert for another opinion on whether AJAX is appropriate for the project. You decide to use [name of expertise location system] to find an expert in AJAX to contact.”*

Due to the geographic spread of participants and to facilitate ease of setup, I conducted this study over the phone. Conversations were recorded with the permission of participants. I felt that telephone interviews were an acceptable research method given that it would not be possible to meet with all my participants face to face. As the participant entered the search term, the researcher would do the same. The way the system operates, anyone typing in the same search term time will see the same results.

Once the results appeared, participants were given time to review the set of names. The researcher then asked which of the 10 experts the participant would like to find more information about. There was no limit on the number of choices. On average, a participant considered finding more information about roughly 3 experts.

After the participant informed the researcher which experts they would like to find more information about, the participant was asked to go to the profile page of each expert they were considering. After visiting a profile page, each participant was told to look carefully over the different information displayed, paying special attention to how helpful the information is in helping them decide to hypothetically contact the person. Once the participant had a chance to look over the profile pages of all the experts she was considering contacting, the researcher would ask the participant to provide a rating on a scale of 1 to 9 (where 1 = not likely at all and 9 = extremely likely) of how likely she was to contact each expert. The researcher would then ask the participant to state in her own words her reasons for hypothetically contacting someone as well as not contacting someone. Finally, the researcher would ask about the number of people in the ‘recommended path’ and ‘alternate path’ since that information is personalized for each user. The steps of the AJAX user scenario is illustrated in Figure 11.

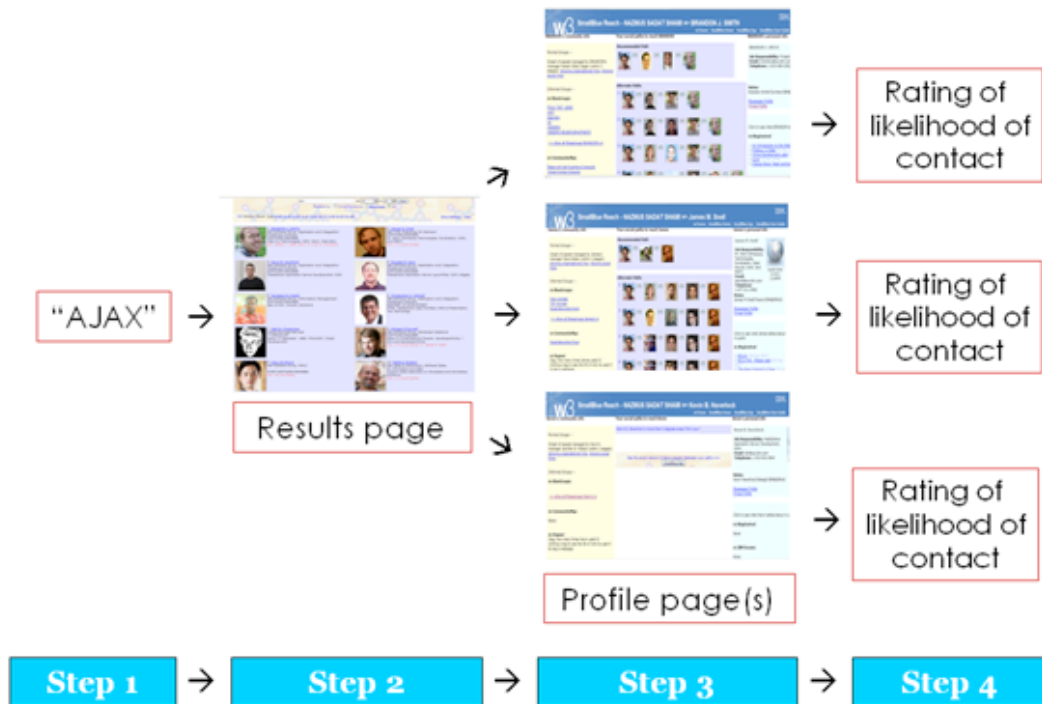


Figure 11. Steps of Phase 2, Part A

### *Why AJAX?*

I chose AJAX as the query term since it was one of the most frequently searched keywords, as obtained from logs of the system. In order to determine the effect of participation in social software as a predictor of considering contacting someone, I needed an expertise keyword that would have data points across different categories of social software. Essentially I needed a keyword that would be blogged about, talked about in forums, and bookmarked and tagged. The AJAX keyword satisfies these criteria in most respects.

Unlike prior studies of searching behavior (Guan & Cutrell, 2007; Pan et al., 2007), I did not use a proxy to manufacture search results. Although the data in the expertise locator system updates and changes dynamically, the same set of 10 names appeared for all my participants.

The list of top ten experts provided us with an interesting dataset to understand the influence of social closeness and participation in social software. Only nine (13.43%) of my participants knew at least one expert directly. The experts also had wide variability in their social software participation. Figure 12 shows the number of social bookmarking tags, blog posts and forum posts of each expert. As can be seen, there is considerable variation among the top ten AJAX experts. In particular, experts in rank 3 and 5 have not participated in social software at all. It should be noted that the system does not filter tags, bookmarks, blog posts or forum posts based on the search term. It merely displays all the information in an unfiltered manner. Expertise rank is also not affected by participation in social software. The system determines expertise solely based on mining email and instant messaging conversations.

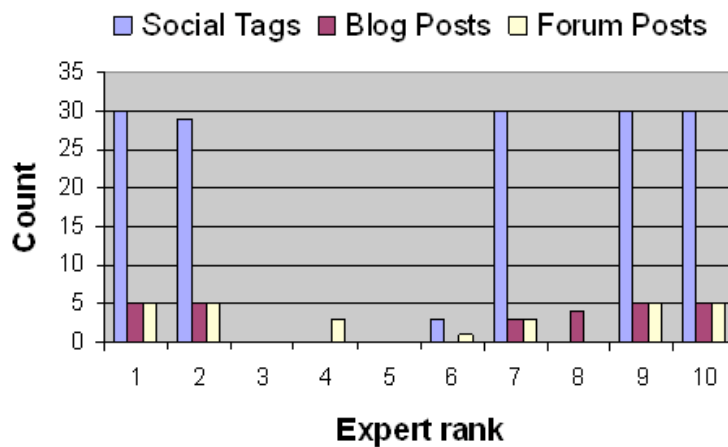


Figure 12. Social software participation of top ten AJAX experts

### *Measures*

I had both quantitative measures that were obtained from SmallBlue as well as qualitative coding that was done on the responses of participants.

## ***Quantitative Measures***

### *Likelihood of considering an expert for contact*

My dependent measure was a continuous variable on a scale of 1 to 9 (1 = not likely at all, 9 = extremely likely) measuring the likelihood of contacting each of the top ten experts that were considered by the participant, as reported by them during the interview. The mean rating was 6.9 ( $SD = 2.69$ ). The expert with the highest rating was considered to be the expert that a participant would hypothetically contact.

### *Expertise*

In the expertise locator tool I used, a search for an expert returned a relevance ranked list. There were 10 experts per page, 2 per row, ordered from left to right and top to bottom resulting in 5 rows of experts per page. I coded expertise as a categorical variable ranging from 1 to 5 corresponding to the row. Previous research found that rank order, expressed by row, was a significant predictor of selecting an expert for further exploration (Shami et al., 2008). I thus considered rank order as a proxy of expertise – the higher the rank order (row 1), the higher the quality of expertise. Rows were dummy coded with row 1 as the base category.

### *Social closeness*

Social closeness was coded as a continuous variable on a scale of 0 to 6 where 0 = know directly and 6 = more than six degrees away. This was obtained by asking participants how many people were in between them and the expert in the recommended path on an expert's profile page. For example, if the participant reported that there were two people in between her and the expert, this was coded as being 3 degrees away. Since the system only displays connections up to six degrees, the lack of a recommended path was coded as the expert being more than six degrees away. This variable was then reverse coded as a measure of closeness. The mean closeness for the experts considered was 2.59 ( $SD = 2.09$ ).

### *Familiarity with AJAX*

Participants were asked to rate their familiarity with AJAX on a scale of 1 to 5 where 1 = I have not heard of AJAX before, and 5 = I use it regularly. The average rating was 3.81 with the majority of participants reporting that they had heard of AJAX but had no training in it. I used familiarity with AJAX as a control variable since I expected people who were more familiar with AJAX would rate experts differently than people who were not familiar with AJAX.

### ***Qualitative Coding***

In order to obtain a better appreciation of the factors that determine likelihood of contact, I coded the reasons participants mentioned for contacting and not contacting a particular person. I derived coding categories by transcribing responses from the audio files and examining the responses. This led to categories related to social closeness/distance, geographic closeness/distance, and inferences about expertise and responsiveness participants could and could not draw from an expert's profile information. I then assigned responses into categories. My committee member Kate Ehrlich and I categorized the set of responses independently. Intercoder reliability using Cohen's Kappa was 0.89 ( $p < 0.001$ ). Disagreements were resolved by discussion. Representative quotes of the categories, as they relate to my hypotheses, are included in the results section. Figure 13 illustrates the reasons participants provided about contacting someone while Figure 14 shows the reasons coded behind not contacting someone.



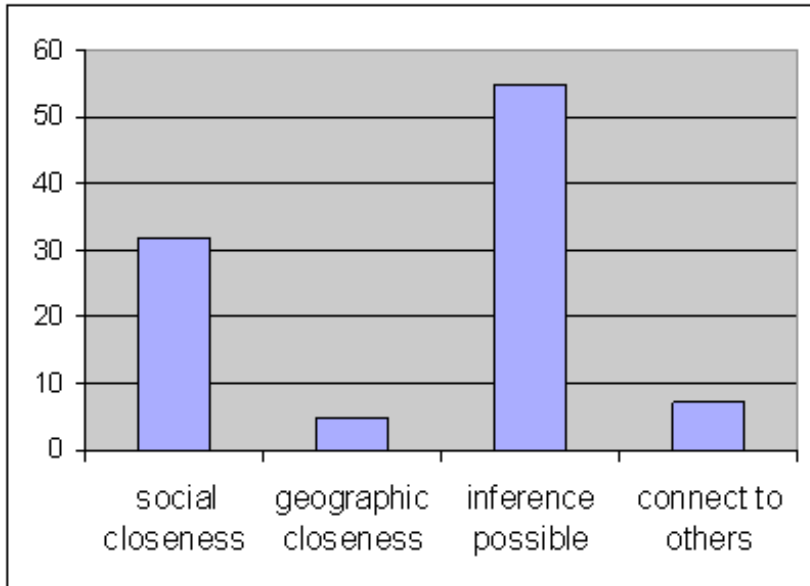


Figure 13. Reasons behind contacting an expert

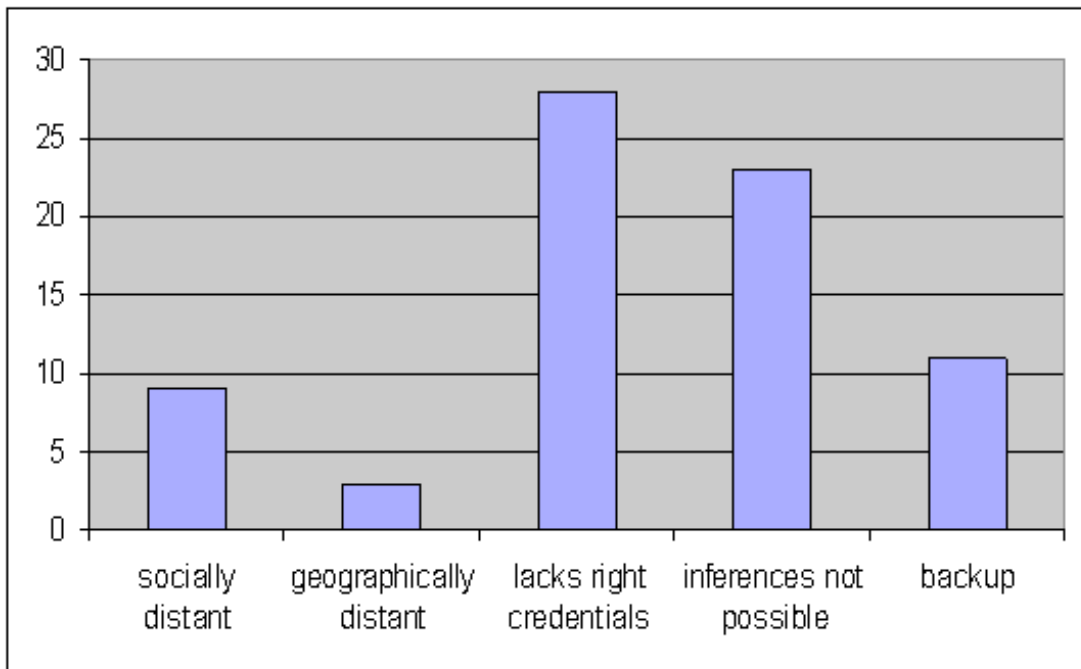


Figure 14. Reasons behind not contacting an expert

## ***Results***

Each of my participants selected three experts, on average, from the initial search page, to gain further information before deciding who to contact. Thus each participant contributed multiple observations which violated the key assumption of independence of observations in multiple regression. To account for this, I ran a multi-level regression model with participant ID entered as a random effect. Results of my analysis are summarized in Table 4. I first entered all my predictor variables into the model. I then removed predictor variables that were non-significant predictors (expertise and AJAX familiarity). While inspecting the scatterplots of the remaining variables, I noticed that the ‘Social software participation’ variable displayed a flattening out pattern. So I entered its quadratic form in addition to its linear form in my model.

Table 4.

Results of multi-level regression model for Phase 2, Part A. Note: \*p < 0.01, \*\*p < 0.001

Parameter	Estimate	SE
<i>Full model</i>		
Intercept	6.55**	1.45
Experts in row 2 vs. row 1	0.02	0.72
Experts in row 3 vs. row 1	0.11	0.65
Experts in row 4 vs. row 1	0	0
Experts in row 5 vs. row 1	0.47	0.6
Social closeness	0.42**	0.11
AJAX familiarity	0.02	0.2
Social software participation	0.002*	0.0005
<i>Reduced model</i>		
Intercept	6.01**	0.56
Social closeness	0.29*	0.1
Social software participation	0.01**	0.002
Social software part. * Social software part.	-5.74E-6*	1.64E-6

### ***Participation in social software***

I found that participation in social software was a significant predictor of likelihood of contact. Hypothesis 3 was thus supported. Posting one more tag, blog, or forum post increases likelihood of contact by 0.01 points. The range of this variable is

0 to 1100 and the co-efficient value is based on the addition of just one more tag, blog, or forum post. A different metric of social software participation (e.g. dividing it by 100) would show a bigger co-efficient value. Importantly, the effect is very significant ( $p < 0.001$ ).

Participants valued social software participation in conjunction with other information in a profile.

*“One because based on descriptions, the whole job responsibilities, descriptions, the blogs, contributions and things like that, yes, that both of them have a fair bit of expertise. Because I see a link between these guys and Tom XX, who I’ve known for a while”*

Having some participation in social software was also seen as compensating for lack of other factors. Thus, one participant said in deciding not to contact a person,

*“This person is more than six degrees away and he doesn't even have a blog”*

Additionally, the quadratic form of this variable shows diminishing returns, indicating that after a certain threshold, participation will not increase likelihood of contact ( $p < 0.01$ ).

I should note again that the tags, blog posts and forum posts displayed were not filtered based on AJAX, but represented the target’s most recent entries on any subject. The significant finding of the mere act of participation, regardless of the content of that participation, as a predictor of likelihood of contact is an interesting and novel finding.

### ***Social closeness***

Social closeness, that is the number of degrees the target was from the participant, was a significant predictor of intent to contact. In other words, respondents rated potential experts higher when those people were within a few degrees rather than further away. Each degree increase in social closeness corresponds to a 0.29 point increase in likelihood of contact ( $p < 0.01$ ). As can be seen from Figure 15, the difference of mean social closeness of experts contacted and those that were considered but not contacted was significant ( $t(49) = -3.08, p < 0.01$ ). These results support hypothesis 4.

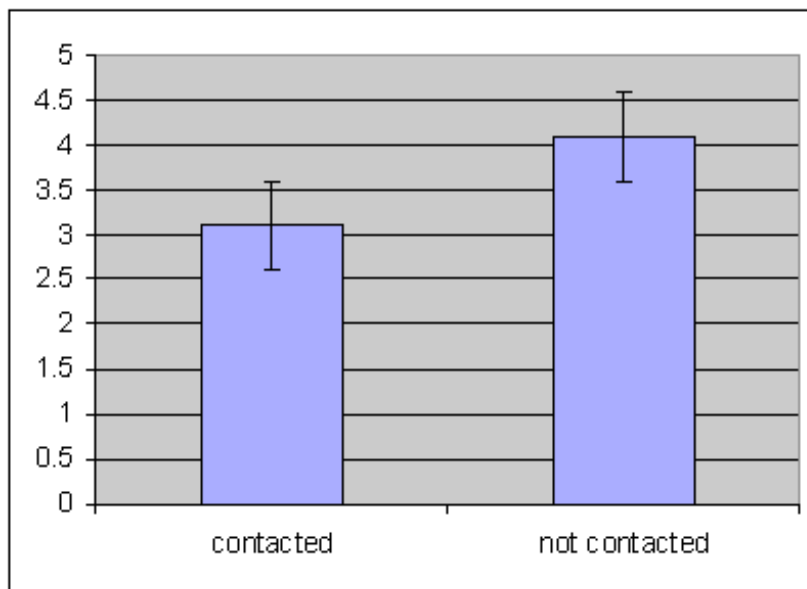


Figure 15. Mean social closeness of AJAX experts contacted and not contacted

The tool I used provided participants with suggested social paths to reach the selected expert and this information was regarded as very helpful in reaching the person.

*“I know the people that the system recommended to go through. If I contact them, I’ll be able to get straight to him”.*

The lack of a path caused others to decide not to contact someone.

*“He is more than 6 degrees away”.*

Participants also valued intermediaries that could help connect to others. One participant said,

*“He’ll be able to help me by passing me to someone that can help”.*

Often participants used the list to identify ‘backup’ people in case their first choice did not respond.

*“If the first person contacted was not available I would just go down the list and contact others”.*

### ***Quality of expertise***

In my analysis I found that expertise, as operationalized by row-based rank order, did not predict the likelihood of contact. Hypothesis 5 was thus not supported. There may be reasons for this hypothesis not being supported. In this phase, I examined the likelihood of contact from amongst a short list of candidates that the participant has already identified. Previous research (Shami et al., 2008) demonstrated that rank order does affect which expert a participant will consider for further examination. In this phase I examined whether remaining differences in rank order

affect likelihood of contact. The lack of significance (as shown by the non-significance of any of the expert rows in table 4) implies that once a user has selected a short list of candidate experts, further differences in rank order have no effect. A summary of hypotheses supported can be found in Table 5.

Table 5. Summary of hypotheses and related results.

Hypothesis	Supported?
H1: Higher expertise rank will be positively related to clicking on a search result for further exploration.	Yes
H2: Existence of social connection information will be positively related to clicking on a search result for further exploration.	Yes
H3: Participation in social software will be positively related to likelihood of considering contacting someone.	Yes
H4: Social closeness will be positively related to the likelihood of considering contacting someone.	Yes
H5: Expertise rank order is positively related to the likelihood of considering contacting someone.	No

The lack of additional effects of rank order is supported by comments from my participants especially one person who summed up the attitude of many others:

*"The best expert isn't the one you're necessarily going to contact."*

Another said:

*"I'd rather have someone who might not be as smart about it, but who knows me really well... I trust him."*

The qualitative analysis revealed that, at least in this setting for the given task, that users were more interested in what could be inferred about the type of knowledge an expert had than about the person's rank ordering.

But the information on the profile page was definitely useful for helping participants make decisions about who to contact. One participant said,

*"He had enough info in his profile that led me to further believe he could help me".*

The information on the profile page could also be used to decide against contacting someone

*"Not as involved in AJAX as I expected"*

Another said,

*"At first looked interesting based on expertise in AJAX widgets, but it's not well documented and don't have much other information to make a decision".*

Figure 16 shows the number of times an expert was considered and then ultimately contacted. As can be seen from the figure, participants considered multiple experts, but after digging deeper into their profiles, eventually settled on a fewer number of experts. A closer inspection of the figure also reveals that the proportion of experts considered and eventually contacted is relatively higher for lower ranked experts than for higher ranked ones.



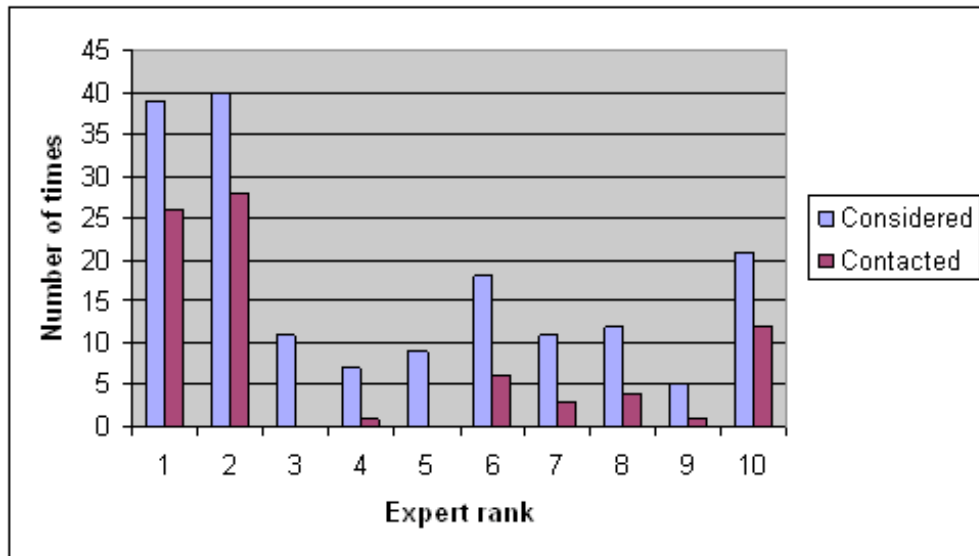


Figure 16. Number of times an expert was considered and contacted in Phase 2, Part A

#### *A note on the role of geography*

In this study, I had participants from 21 countries. I wanted to take advantage of this fact by ascertaining whether geography could be considered a signal that influences likelihood of contact. However, intuitively one would think that social closeness and geography might be related and confound my findings. That is why I initially performed my data analysis by excluding geography. Afterwards I was still intrigued by the role of geography and decided to include it in my model. Low intra-class correlation ( $< 4\%$ ) in the multi-level model illustrated in Table 4 can be used as justification to run my model using regular regression. Regular regression did not show any multicollinearity, suggesting social closeness and geography were not related in my data. I also ran my multi-level model with only North American subjects rating only US experts, and my results still held at the  $p < 0.10$  level. If geography were confounding social closeness, significant differences would not have been

obtained. I will thus discuss how geography could be an important signal in determining whom a person contacts for specific expertise.

Despite advances in information and communication technology, geographic distance has shown to provide strong challenges for people intending to share knowledge (Olson & Olson, 2000). Allen has proposed that when people are apart more than 30 meters their interaction is negatively impacted (Allen, 1977).

In order to understand the barriers of distance, the considerable benefits of contacting others in closer proximity needs to be understood. Individuals in closer proximity benefit from easy access to each other through shared time zones, culture and even language, which lowers communication costs. This may create common ground for communication (Clark, 1996). When located in the same office, colleagues can ‘bump into’ each other, which serves as a reminder of things promised but not delivered. They can see when others are available. Herbsleb, Mockus, Finholt & Grinter studied software engineers located in the U.K., Germany, and India, as they collaborated on integrated and time-sensitive software development projects (Herbsleb, Mockus, Finholt, & Grinter, 2000). They found that requests for modifications in software took longer whenever they involved engineers in multiple locations. These engineers also reported sharing less personal information and having less ‘affective trust’ (McAllister, 1995) with their distant colleagues.

It’s not that individuals farther away in geographic distance intentionally ignore requests from distant colleagues. When forced to make difficult allocation choices, the social pressure and multiple awareness cues of closer proximity may overwhelm the relatively sparse communication channels of distant colleagues. In an experimental study, Fussell, Kiesler, Setlock Scupelli and Weisband (2004) found that individuals had difficulty managing time and attention equitably across projects with different geographic configurations. When involved in both collocated and distributed

collaborations, participants favored tasks with collocated partners despite equal importance of tasks. The unequal distribution of attention may partially explain Herbsleb et al's (2000) field findings that software modification requests that originate locally are completed more quickly than distant ones. Finally, simulations of distributed work have found that workers at the same sites formed strong in-groups, and enlisted help from collocated colleagues at a much higher rate than from remote colleagues (Bos, Shami, Olson, Cheshin, & Nan, 2004; Shami et al., 2004). The strong local in-groups inhibited cross-site collaboration and resource exchange. Taken together, all these findings point toward the considerable advantage of contacting others closer in geographic distance. I can thus hypothesize that geographic distance will be negatively related to the likelihood of considering contacting someone.

Accordingly, I created a categorical variable with 4 levels representing the 4 geographic locations (US, India, China, France) of the AJAX experts. This variable was then dummy coded with the US as the base category.

I ran a multi-level regression with the same variables as in Table 4. The graphs of 'familiarity with AJAX' and 'social software participation' variables displayed a flattening out pattern so I entered their quadratic forms in addition to their linear forms in my model. I found partial support for the hypothesis that geographic distance will be negatively related to the likelihood of considering contacting someone. Being in India and China was negatively related to likelihood of contact, while being in France had no effect. Since the majority of my participants were from the United States, they perceived communicating with a person in France less of a communication cost because of linguistic and cultural barriers than with someone in India or China. I should note that this result should be taken with a grain of salt since *all* participants were not from the Western hemisphere. However, since the majority

of participants were from the Western hemisphere, and my findings still hold after adding geography to the model, geographic location could be considered a signal that influences expertise seeking behavior. Table 6 illustrates the change in my results once I include geography.

Table 6.  
Results of multi-level regression model for Phase 2, Part A, when geography is included. Only significant predictors shown. Note: \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001

Parameter	Estimate	SE
Intercept	9.13***	0.16
Social closeness	0.28*	0.09
US and India	- 3.71***	0.82
US and China	- 2.89**	1.08
AJAX familiarity	- 1.94*	0.80
AJAX familiarity * AJAX familiarity	0.29*	0.13
Social software participation	0.01***	0.002
Social software participation * Social software participation	-6E-006***	1.8E-006

### **Phase 2, Part B: The role of individual pieces of information within a profile**

Given the plethora of signals available on the SmallBlue profile page of a user, I wanted to determine the individual contribution of different pieces of information as signals. Figure 8 shows a screenshot of a profile page. As can be seen from the screenshot, there is a lot of information available in the profile. Some information within the profile could be considered as ‘assessment signals’ since they require possessing the quality being signaled. Profile information that fall into this category

are social connection information, basic corporate directory information, and mailing list membership.

The system calculates social connection information based on actual communication. This prevents artificial inflation of one's social network connections. Social network connection information inferred through actual communication is thus a piece of information within the profile that is hard to fake. The system displays 15 social connection paths from the user to the expert, up to six degrees away. The path displayed at the very top is considered to be the 'recommended path' or the path from the user to the expert that is shortest, as well as strongest based on communication patterns. The system displays the remaining connection paths in descending order of tie strength.

Basic corporate directory information includes a person's job title, job description, and geographic location. This information is entered automatically for every employee by the organization, leaving no room for deception. Basic corporate directory information is displayed on the top right hand side of the profile page. Within the organization I studied, individuals were subscribed to certain mailing lists by the organization itself. The basis for this auto-subscription was a determination by the organization that the employee needed to belong to the mailing list based on their particular business unit or skill-set. Employees could self-subscribe to mailing lists as well, but the majority of profiles that my participants looked at did not have many self-subscribed mailing list membership. Mailing list membership is displayed on the top left hand side of the profile.

The profile also contained pieces of information that were user-generated content and could be utilized for self-presentation. These included social tags and bookmarks, blog posts, forum posts, and self described expertise. The bottom left hand side of figure 1 shows the top 30 social bookmarking tags of a user, followed by the

number of times the tag has been used. On the right hand side of the page, the 5 most recent blog posts and their timestamp, the 5 most recent forum posts and their timestamp, and the 5 most recent bookmarks and their timestamp are displayed. The timestamps provide an indication of the recent activity level of the expert. Below the social bookmarks is an area where employees can describe their skills and the projects they've worked on.

The different pieces of information within a profile could be considered to represent behavioral, social and personal characteristics of any expert the user is considering contacting. It is worth mentioning that the data aggregated together by the system presents information "as is" from those sources. There was no attempt to aggregate the different elements into any kind of metric or weight any one element differently from any other nor is there any editing of the elements except to limit the number of entries in any one category to fit in the available space.

### ***Procedure***

For this part of my study, I had my participants imagine themselves in the following scenario.

*"I want you to reflect back on a situation during your career at [company name] where you needed to locate people that have expertise on a certain topic. I'll give you some time to think about this expertise. Once you've thought about it, let me know the expertise keywords you would use to search for a person with that expertise."*

The researcher would then ask the participant to provide a rating on a scale of 1 to 9 (where 1 = not important at all and 9 = extremely important) regarding how important it was for the participant to find the right person to contact. The mean rating

on this scale was 8.08 (SD = 1.47), indicating the high importance of finding the right expert. Figure 17 shows the steps in the retrospective reflective scenario.



Figure 17. Steps of Phase 2, Part B

Participants were then told to enter the keyword they would use to search for a person with the expertise they sought. This is the first step of the user scenario depicted in Figure 17. Expertise keywords entered by participants had a wide variety, but were mostly related to technology.

Once the results appeared, participants were given time to review the set of names. The researcher then asked which of the 10 experts the participant would like to find more information about. There was no limit on the number of choices. On average, a participant considered finding more information about roughly 3 experts. After the participant informed the researcher which experts they would like to find more information about, the participant was asked to go to the profile page of each

expert they were considering. After visiting a profile page, each participant was told to look carefully over the different information displayed, paying special attention to how helpful the information is in helping them decide to hypothetically contact the person. After a participant told the researcher that she was finished looking over all the information in the profile, the researcher would ask the participant to provide a rating on a scale of 1 to 9 (where 1= not helpful at all and 9 = extremely helpful) how helpful each of 7 pieces of information (mailing list membership, social tags and bookmarks, social connection paths, basic corporate directory information, blog posts, forum posts, and self described expertise in the corporate directory) were in helping her to decide whom to hypothetically contact for the expertise she sought. Often when providing ratings participants would spontaneously justify the reasons behind their ratings. Occasionally the researcher would probe participants when they provided particularly high or low ratings. Once the participant had a chance to look over the profile pages of all the experts she was considering contacting, the researcher would ask the participant to provide a rating on a scale of 1 to 9 (where 1 = not likely at all and 9 = extremely likely) of how likely the participant was to contact each expert. All steps of this user scenario are illustrated in Figure 17.

### ***Qualitative analysis***

In order to obtain a grounded appreciation of the people sensemaking process, I completely transcribed all audio interviews. I then carefully read through the documents highlighting parts that were related to perceptions of the 7 pieces of profile information I was considering. I then organized the relevant parts into common themes, and coded the documents using the themes that emerged. Representative quotes from these themes in relation to the 7 pieces of information are included below.



## **Results**

Each of my participants looked at roughly three profiles before deciding who to contact. Thus each participant contributed multiple observations to my dataset, which violated the key assumption of independence of observations in multiple regression. To account for this, I ran a multi-level regression model with participant ID entered as a random effect. Results of my analysis are summarized in table 7. In the following, I discuss the statistical results of each piece of information I was interested in, and how my participants interpreted the digital artifacts to infer expertise, access to the expertise and likely responsiveness of the expert. I use concepts from signaling theory in orienting the discussion.

Table 7.

Results of multi-level regression model for Phase 2, Part B. Note: \*p < 0.05

<b>Parameter</b>	<b>Estimate</b>	<b>SE</b>
Intercept	0.26	3.1
Corporate directory	0.36*	0.17
Blog posts	-0.04	0.21
Forum posts	-0.05	0.29
Self described expertise in corporate directory	- 0.09	0.22
Mailing list membership	0.36*	0.18
Social network connections	0.37*	0.17
Social tags and bookmarks	0.13	0.18

### ***Mailing list membership information***

For each unit increase in the perceived helpfulness of mailing list information, likelihood of contact increased by 0.36 points ( $p < 0.05$ ). The helpfulness of mailing list information was primarily related to being members of the same mailing list which created common ground, as well as being familiar with the mailing lists listed on an expert's profile. This provided a reliable signal of expertise.

*"It's helpful because you know where he is active."*

Another participant mentioned the importance of being on a relevant mailing list.

*"because he is a member of the design and user experience community... which is pertinent to the question here"*

### ***Social tagging and bookmarking information***

Perceived helpfulness of social tags and bookmarks were not found to be a significant predictor of likelihood of contact ( $p = 0.46$ ). In the words of one of my participants:

*"For me personally the problem is I am not using [social bookmarking], so from that perspective this doesn't really give me a good indication on how useful are these kind of things and in this specific context GPFS has nothing really to do with these kind of offerings."*

Yet another participant responded:

*“I’m mixed emotions about social bookmarking. People don’t necessarily tag or bookmark stuff about stuff they know. I most of the stuff that I tag is stuff that I’m trying to learn. So would I go to that person for expertise?”*

Some pointed to the novelty of the technology and its lack of use within the organization.

*“I’m kind of neutral on tags because in the great scheme of things it’s relatively new and I don’t think a lot of people just gut feel are not using it on any kind of regular or productive basis.”*

### ***Social network connection paths***

Social network connection paths were significantly helpful in assisting a user to decide whom to contact ( $p < 0.05$ ). Out of all the information available in a profile, perhaps social network connection information could be considered the strongest assessment signal. This signal was primarily used to infer accessibility through details of social paths

*“...it wouldn’t be too much of a cold call to say ‘hi, I understand you know my colleague so and so, I’m calling you about this other topic.’ I guess it would make me feel more comfortable knowing that I could sort of name drop.”*

Social connection paths also were strong signals of expertise since an expert would be linked to other experts within a connection chain.

*“Looking at the alternate paths, you get credentials this is clearly someone who, as I look at the alternate paths, there are like a ton of people that you know he's one step away from, that further credential him”.*

Another participant mentioned:

*“There's two things I learned from the alternate path. One is, that he's one step away from me by two people that I work with all the time, and I trust their judgment”*

### ***Corporate directory information***

For each point increase in helpfulness of corporate directory information within a particular expert's profile, the likelihood of contacting that expert increases by 0.36 points ( $p < 0.05$ ). Directory information provided key summary information such as job title and responsibility, which were reliable signals of expertise.

*“HR folks talk about looking at resumes and getting an impression of a person in the first five minutes. So in a social networking environment or something like this, I want it to be even faster and I want to have, there's got to be something there in the first thirty seconds that catches my eye, that's going to draw me to that person to look up more information on them. And with this person's title description, role, what he is in the company, that's exactly what I was looking for”.*

Directory information also conveys the seniority of an employee, which in the case of this participant, influenced him not to contact the expert.

*“Also you know if it's worth my time to reach out to him. This person is you know reports to an executive vice president within [the company name] so obviously being that high up in the organization you know probably it's not worth his time to respond to me and it's not worth me you know my time you know simply because chances are my email is going to get lost or however I choose to contact him is gonna be you know...”*

### ***Blog posts***

Although a convenient platform for self presentation, perhaps the fact that anyone can blog led to the non-significance of blog post information ( $p = 0.84$ ).

*“People who blog are people who. .. have a lot of time to talk about it. Anybody that I know, a deep subject matter expert, rarely has the time to talk about it”*.

### ***Forum posts***

Similar to blog posts, forum posts could be used for self-presentation, but were not found to be significantly helpful to my participants ( $p = 0.83$ ). Although forum posts did not turn out to be statistically significant, through qualitative coding I found that many participants viewed experts' forum posts as their willingness to respond to unsolicited queries.

*“I see that this person is involved in... in forums, and so on. I see that this person is quite open to contact. I will feel free to just contact him directly.”*

### ***Self described expertise in corporate directory***

Self descriptions of expertise perhaps provided the most opportunity for self presentation among the information sources I looked at. But similar to other sources that are user generated, information from this source was not statistically significant ( $p = 0.68$ ). Users concerns with self reported expertise involved listing too many skills and projects that one could conceivably be an expert in or involved with, as well as the lack of updates of such information.

*“My reservation is data quality in [name of corporate directory] is sometimes questionable. If information on [name of corporate directory] would be reliable and if the information if the people manager would push people to fill out the [name of corporate directory] information correctly I would find it useful. But currently I don't find it very useful”.*

### **The role of participation in social software as a proxy of approachability**

I was intrigued by the finding that participation in social software mattered whereas expertise rank did not. To understand this better, I conducted 18 follow-up interviews where I asked participants about their attitudes towards people that participate in social software. When I mentioned to one participant that participation in social software such as social bookmarking and tagging might be considered an indicator of *interest* rather than *expertise*, her response was:

*“My assumption is that if you're interested in it, you probably know something about it. I assume that there are multiple experts out there in varying degrees and I might not need the grand daddy of them all expert”.*

On the subject of why there is a perception that participants active in social software are more likely to respond to an expertise request, one participant mentioned:

*“I see that this person is involved in [social bookmarking] tagging and in forums, and so on. I see that this person is quite open to contact. I will feel free to just contact him directly”.*

Yet another participant said:

*“People who use [social bookmarking] or forums are more likely to reach out to the community with their questions and their expertise and therefore I would think they would be more likely to assist in sharing their own expertise.”*

The social bookmarking also provided an additional avenue into expertise:

*“More information about AJAX was being referred to in his expertise profile and [name of social bookmarking software] inferring that he works on that as part of his daily job”.*

Conversely, I found that the lack of information led participants to be less interested in contacting the person. One person gave as part of their reason for not contacting:

*“No blog or forum entries”*

It appears that individuals that participate in social software are perceived by others to be creating social capital. Adler & Kwon refer to social capital as the goodwill engendered by social relations that can be mobilized to facilitate action (Adler & Kwon, 2002). They contend that if goodwill is the substance of social capital, its effects flow from the information such goodwill makes available. For instance, one participant responded:

*“Once I find somebody, I need to find out first of all what is, how competent are they. And second of all how benevolent are they. The act of them sharing gives them a lot of points in my book because it tells me they’re willing to um help.”*

Interestingly, creating goodwill reflects findings of motivations behind participation in user generated content such as social software pretty well. In a study of Wikipedia contributors, it was found that altruism and benefit to the community were primary motivations for contribution (Oreg & Nov, In Press). This dissertation lends support to the idea that the same perceptions of altruism might apply to people who actively participate in public forums such as blogs, wikis and social bookmarking systems. In organizational settings, employees are not paid to blog or participate in forums, and the opportunity cost of such participation leaves employees less time to focus on their primary task. Yet through such participation, individuals may be signaling that they are more efficient with their time and have the greater good of the community in mind. In this research expertise rank order was not significant. This suggests that expertise is a necessary but not sufficient condition for likelihood of contact. My participants felt that those who were already sharing their knowledge through social software participation are more likely to respond if contacted.



## CHAPTER 5

### CONCLUSION

#### **Bringing the pieces together**

This research sought to examine the process through which individuals go about considering contacting others for assistance with accomplish non-routine, complex work. It commenced with the premise that the widespread popularity of social computing tools and increased growth and availability of expertise locator tools would assist them in the expertise location process. Since individuals usually use these tools only after they have exhausted their personal network of contacts (Borgatti & Cross, 2003; Cross & Sproull, 2004; Hertzum & Pejtersen, 2000), the search results returned by these tools contain names and related profiles of mostly unknown others. This presents a challenge in evaluating the credibility and suitability of their expertise, and, also assessing the likelihood that a request for information to the stranger will get a response (Ehrlich et al., 2007; Shami, Yuan, Cosley, Xia, & Gay, 2007). No matter how much time and energy we spend gathering information, most choices must be made without complete knowledge about the relevant alternatives. When gauging the expertise of unknown others, the seeker is in a situation of imperfect information. He or she is unsure of an expert's capabilities. Potential experts have observable characteristics and attributes such as previous work experience, education, gender, and race. Unalterable attributes, such as gender and race, are called "indices." Alterable or changeable items, such as education, work experience, and other qualifications, are called "signals" and can be manipulated (Spence, 1973). Signaling theory, originally developed in economics (Spence, 1973) and biology (Zahavi, 1975), can be used as a theoretical framework to explain how information from digital artifacts can be used to form impressions of credibility, expertise, availability and responsiveness. Because of the expertise seeker's uncertainty about an expert, he or she must rely on signals either

intentionally or unintentionally sent by experts in relation to these qualities. The seeker can then use these signals to draw inferences about the qualities sought by the seeker.

This chapter offers an overview of the findings discussed in Chapters 4 and connects them back to the conceptual framework introduced in chapter 2. It further discusses contributions and implications of the findings. Finally, limitations and potential avenues for related research are considered.

### **Overview of findings**

The problem of seeking to contact others for expertise using technology was approached through the use of a novel expertise locator system called SmallBlue (Ehrlich et al., 2007; Lin et al., 2008). This system was chosen because it was particularly amenable for use in disambiguating the different steps in the technology mediated expertise search process such as generating a query, searching for relevant information, evaluating and making sense of the information found, and coherently integrating different pieces information into a coherent whole to arrive at a decision. It aggregates together widgets of popular social computing technologies such as blog posts, social tags and bookmarks, forum posts, and social network connection information to create a composite profile of an expert. Individuals get to a profile by first entering an expertise search keyword into SmallBlue, looking over initial search result set(s) or a Master page, and clicking on individual search results of interest that take them to a profile or Detail page.

The first set of findings in phase 1, as discussed in chapter 4, pertains to patterns of behavior related to evaluating an initial search result page that contains summary information about experts on the keyword being searched. Part A of phase 1 looked at a single keyword and part B looked at behavior across multiple keywords. The pattern of results held across both studies. The second set of findings of part A

and part B of phase 2, pertains to the factors that influenced likelihood of contact after viewing more detailed information from an expert's profile page. These factors were helpful in explaining the decision process participant's went undertook regarding whom to contact.

In phase 1, which focused on the search result page, there was summary information displayed about each expert. This included an expert's name, picture, business unit, and job description. An additional piece of information, which was personalized to each participant, was a referral chain, upto three degrees away, signaling how a participant could connect to an expert. This social connection information could be considered an assessment signal, since it was inferred through email and instant messaging communication, rather than self-reported. It certainly satisfies the criteria of being hard to fake since it will only show up when a participant has actually had communication above a certain threshold to justify the connection. My results indicate that this snippet of information was a strong predictor of which individual search result a participant would click on for further exploration. In part A, I had a comparatively controlled experimental setup where participants looked at a fixed set of 10 experts related to the query term AJAX. Social network connection information had a strong influence on clicking behavior. So did rank order. Rank order could also be considered an assessment signal since expertise is inferred through the system rather than self-reported. There was no correlation between the display of social connection information and rank order. This meant that a higher ranked search result could have no social connection information and a lower ranked search result could have social connection information. My findings suggest that a user is likely to click *both* the higher ranked result with no connection as well as the lower ranked result with social connection information. While prior studies have suggested social network data could be helpful in expertise search (Ehrlich, 2003; Ehrlich et al., 2007;

Terveen & McDonald, 2005), ours is the first to empirically demonstrate the value of this information. It also extends findings of prior research on rank order when searching for documents (e.g. Granka et al., 2004; Joachims et al., 2005; Joachims et al., 2007; Pan et al., 2007) to the domain of searching for people.

Part B of phase 1 was designed to determine if my findings hold when the expertise keyword used to query the system varies, resulting in a different list of experts for a given query. I found that findings in part A held in part B as well, namely social connection information and rank order were significant predictors of which link a participant decides to click on for further exploration. Again, there was no correlation between social connection information and rank order, suggesting that both these pieces of information were salient as assessment signals that influence clicking behavior.

Phase 2 focused on further analysis of technology mediated expertise location through investigating user behavior around an expert's profile information. The profile contained a wealth of signals, some conventional and some assessment. Un-editable information on a profile page such as job description, geographic location, and mailing lists that the organization subscribes an employee to could be considered assessment signals. Inferred information such as social network connection information and expertise rank order could also be included in that category of signals. Self authored information such as blogs, social tags and bookmarks, and forum posts could be considered closer to conventional signals. My results revealed that participants' expertise seeking behavior was shaped by a number of factors associated with these different types of signals. First, a participant's perception of an expert's participation in social software such as blogs, social tagging and bookmarking, and online forums, regardless of the content of that participation, acted as a proxy of availability and approachability. Participants felt that experts high in social software

participation were sending a signal that they were making themselves available with the aim of creating social capital. Second, the inferences made possible through social network connection information provided further signals of approachability. It is not enough to find an expert; the system must find an expert who is likely to respond. It is fruitless to find the best expert in a subject domain and not get a response to a request for expertise. Social network connections provided possible social conduits that can facilitate the expertise exchange, making a referral chain more likely to succeed. My results demonstrate that participants felt that having a social conduit between them and an expert increased their perception of receiving a response. They also felt that the referral chain often times displayed that an expert is socially connected to other experts in the field. Such social network ties credential an expert since only individuals in high status will have connections to others in similar status. This hard to fake signal increases the credibility of the expert, making the signal more reliable. Third, expertise rank order was not found to be a significant predictor of likelihood of contact. This is in contrast with the findings discussed in part A and B of phase 1. Phase 1 concerned the initial search result page, which only provided summary information about an expert. This was adequate to spark interest and influence clicking behavior to find out more information about an expert. Once participants were able to view the profile, they were presented with a plethora of information. The abundance of information allowed participants to make a more informed decision regarding which expert to contact. This leads to my conclusion that expertise is a necessary but not sufficient condition within expertise seeking behavior. All of the profiles viewed by participants were of experts, although their expertise did vary, as inferred by the system. It is thus evident that there is more than one “expert” that suffices for a given expertise query. In this study, everyone performed the same search which generated a list of the same 10 experts. Yet, there was a lot of variability

in which of the 10 people were considered at all and which ones were likely to be contacted. This finding strongly implies that the person who was regarded as an “expert” for one person was not the same as the “expert” for another. I could say that expertise is in the eye of the beholder.

Overall, phase 1 and 2 supports the assertion that the relationship between the seeker and expert is a salient factor in deciding whom to contact. I explored two aspects of interpersonal relationships: a) social closeness as defined by the number of people on the path from seeker to expert and b) geographical distance. Seekers were more likely to want to contact someone who was closer, socially and geographically than someone who was further away. Several expertise locator systems acknowledge the importance of social closeness as a factor in recommending experts (Terveen & McDonald, 2005) or as in the case of the system used in this study, as an important element in the seeker’s decision process (Ehrlich et al., 2007).

In combination, the phase 1 and phase 2 illustrate the strong influence of social network connection information as an assessment signal influencing expertise seeking behavior. The factor that remained significant across both phases was social connection information. In the studies in phase 1 social connection information signaled social conduits that could facilitate the expertise exchange, increasing the likelihood of response. In phase 2, the social connection referral chain credentialed an expert by illustrating how they may be connected to other experts in the field, increasing the credibility of the expert. Even in the response to the interview question ‘how helpful was this piece of information in helping you to decide who to contact?’, social network connection information came out statistically significant. Taken together, these findings empirically demonstrate the importance of social network connection information in influencing technology mediated expertise seeking

behavior, and the value of displaying such information in systems designed to facilitate expertise seeking.

### **Limitations**

Research into the complex and thus far still relatively unexplored domain of technology mediated expertise search is obviously subject to a number of limitations. Some of these limitations were justified and/or addressed as far as possible in the design of the study itself. Others are acknowledged here in that they suggest avenues for future research.

My choice of studying a single organization might be seen as a limitation since a small sample can be problematic from the standpoint of research generalizability. Can the patterns of behavior identified in the large distributed organization I studied reasonably be considered representative of those in other organizations in other industries? Perhaps not entirely.

The single organization, however, offered both practical and theoretical advantages that were considered to offset concerns regarding generalizability – particularly as the intent of this research was pseudo exploratory rather than theory testing. First, the focus on a single organization enabled a depth of field access that would not have been possible had I tried to split my attention across a number of different companies. Furthermore, by selecting a real world organization, rather than conduct my study in a lab setting, claims of ecological validity could be made.

Another limitation of the research approach was its reliance on data acquired from retrospective data collection techniques in the reflective scenario described in part B of phase 2 in chapter 4. This may be susceptible to biases and rationalizations after the fact. The data collection procedure in this scenario also suffers from some limitations. Participants in this scenario were asked to go through each of the information sources in turn and “on a scale of 1 to 9 where 1 is least helpful and 9 is

most helpful, how helpful was [name of piece of information] in helping you decide whether or not to contact this person.” The rating did not capture whether the respondent perceived the information as positive or negative about the candidate. An example of a negative perception,

*“wow he is very active in [name of mailing lists]. well I’ll tell you what with such a huge list of groups I’m not interested in this person... because he seems to get involved in everything and to me that is contrary to real expertise of a subject matter that would not be the type of individual I would consult.”*

There is some risk of reducing the significance of my results in those cases where the information is helpful but negative. Mitigating this risk is that most perceptions were positive. In the few cases where there was a negative perception, the respondent also gave lower ratings. For instance, in the quote given earlier, the respondent rated the helpfulness of the information as 2.

The meaning of the term “sensemaking” as used in the current research must also be addressed as a limitation. As discussed in chapter 2, there is much confusion about this term and this dissertation did not address or aim to clarify the theoretical debate around this term.

### **Contributions to theory and practice**

This research on expertise seeking behavior using information retrieval systems contributes both theoretical insights and empirical findings relevant to a number of fields of study.

#### ***Information search theory***

First, this research adds to the growing literature and empirical body of work on information search. It takes a different tack to most information search studies that focus primarily on finding documents (e.g. Granka et al., 2004; Joachims et al., 2005;



Joachims et al., 2007; Pan et al., 2007). Instead, it approaches the information search problem from the vantage point of searching for people.

Existing theoretical work on information search has been on how individuals look for documents. For example, Marcia Bates describes the Berrypicking model of information search (Bates, 1989). In this model, information search is akin to picking berries from bushes. Using this metaphor, Bates describes how the right berries to choose are scattered across different bushes. These berries have to be picked one at a time from different bushes. Similarly users usually start with just one relevant reference and move through a variety of sources, each new piece of search result providing a new conception of the search query. At each stage a user modifies both the query terms as well as the search requirements. A decade later, Peter Pirolli and Stuart Card introduced the ‘Information Foraging’ theory of document search (Pirolli & Card, 1999). Using another metaphor from nature, Pirolli & Card draw on ‘Optimal Foraging’ theory in animals (Stephens & Krebs, 1987) in their theoretical formulation. As in the real world where animals forage for food, the online world of the Web is a patchy environment with useful information arranged in different clusters. Patches of useful information reside in different websites, and as “informavores” humans seek out the richest patches and extract useful information. As humans forage for more information, it becomes harder to find additional useful information from the same patch. Such diminishing returns cause humans to ‘feed’ at a patch until the rate of gain of useful information falls below the perceived average. Once it is thought that the grass is greener in another ‘patch,’ information seekers switch to another page or website, or reformulate their search query, seeking out more fruitful patches of information.

To a large extent, these theories were formulated when searching for documents was the prominent paradigm of information retrieval. Recently, we are

witnessing an increase in the need to search for people. This is evident from the use of ‘Googling someone’ in everyday vernacular, to more specific applications of finding others with similar interests in social networking sites, or finding a romantic partner in online dating sites. Although document search and people search share similarities in both being an information retrieval problem, there are reasons to believe that searching for people differs from searching for documents in significant ways.

Relevance is important for both document search and people search. It is crucial that the most relevant results based on a query term are displayed in the first few results. An element of relevance is credibility and trust of the source, which is again, important for both document search and people search. However, for people search a relevant result is different for different people, above and beyond the degree it is for document search. There are critical social factors pertinent to people search that set it apart from document search. Depending on the goal users have when they search for people, a result that is relevant for one person may not be relevant for another person. If one seeks to obtain diverse knowledge not found within one’s own social network, an unknown expert might be relevant (Granovetter, 1973). On the other hand, if one seeks to obtain tacit knowledge which may require multiple iterations of back and forth, someone socially close might be relevant (Hansen, 1999). In the context of searching for someone to contact to ask for advice or expertise, one needs to take into consideration the knowledge seeker, the knowledge source, and the relationship between the two (Cross & Sproull, 2004). Factors such as familiarity with a person (Fidel & Green, 2004), accessibility, responsiveness, the ability to receive a response in an understandable manner without being constrained by barriers of language and culture, the respondent’s ability to express tacit knowledge, the opportunity to have an interactive dialog where concerns can be addressed over multiple interactions, and affective dimensions such as comfort level with a person

need to be taken into account. In document search, these factors do not come into play. A user can judge whether a document is relevant or not by reading through it. They need not worry about the relational factors mentioned above.

### ***Signaling theory***

Second, this dissertation contributes conceptually to signaling theory by developing an interpretive framework for analyzing the decision processes involved in ‘people sensemaking’, which integrates prior insights from multiple streams of research to characterize expertise seeking practices in a real world work context.

Let me revisit my preliminary conceptual model of people sensemaking in expertise seeking behavior, as shown in Figure 18.

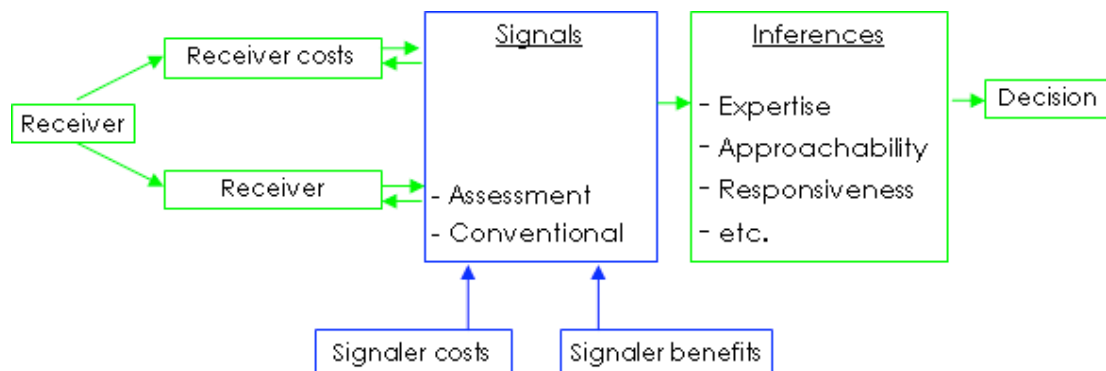


Figure 18. A preliminary model of 'people sensemaking' in expertise seeking behavior - revisited

The data in this research reveal the nuances of expertise search by which individuals successfully make decisions under uncertainty to accomplish the complex task of finding someone to contact for specific expertise, thus contributing to a better understanding of how people make sense of digital information about individuals they do not know. Empirical contributions include the application of signaling theory to a new context of human communication. By explaining individuals' choices of signals within digital artifacts that they considered influential, this research offers new

understanding into how a theory used primarily in biology and economics can provide explanatory power to pieces of information within technologies designed to augment and assist the expertise location process.

In the different parts of phase 1 described in chapter 4, receivers of signals quickly gravitated to signals they could easily interpret. A familiar name displayed in a social network referral chain quickly grabs attention, and thus influenced clicking behavior. Rank order is also easily interpreted. It appears that the use of search engines have conditioned us to the “I’m feeling lucky”<sup>10</sup> effect – higher ranked results carry more trust than lower ranked results. In phase 2, we again see that familiar names within a social network connection chain are easily recognizable and perceived as adding credibility to an expert. Receivers also value information that they feel is hard to fake, such as high participation in various forms of social software, as well as job descriptions which cannot be edited.

Signaling theory was also used as a way of differentiating the value of different sources of information. As its central premise, signaling theory holds that information which is hard to fake is more reliable. While there is an abundance of conventional signals online, my study reported results from an online profile that had both conventional and assessment signals. In looking at these signals within the digital realm, I found that content that is directly user generated such as blogs, social bookmarks, tags, and self-described expertise could be open to manipulation and was not considered helpful in helping to select an expert. Conversely, information that comes from mining data whether directly from a corporate personnel database or by inferring from communication records as in the case of social network information, is less open to direct manipulation.

---

<sup>10</sup> “I’m feeling lucky” refers to the search button on the popular search engine Google, which when clicked takes the user directly to the top ranked result of a given query.

However, where signaling theory makes a categorical distinction between cues that are hard to fake and those which are easy to fake, in the digital realm, cues really fall on a continuum from those which are entirely user generated to those which are entirely mined from data over which the user has no direct control. Thus, I consider tags and self-rated expertise to represent information over which the user has complete control whereas the user has no direct control over corporate HR information. For example, social network connection information is a more reliable signal of expertise because people within a social network connection chain can credential the expertise of an individual. A middle ground is occupied by information such as mailing list membership. While it is unlikely that people will join a community with the intent of making their interests visible to others, membership in a community is under an individual's control.

Within the digital realm, we can thus think of signals falling along a continuum of conventional signals and assessment signals. Mined sources of information such as social connection information and expertise rank order could be considered to be more along the assessment signal end of the continuum while self-authored data such as blog posts, social tags and bookmarks, and forum posts could be considered along the conventional signal end. Mailing list membership could be considered to occupy a middle ground.

Given the finding that my participants relied primarily on social network data, community membership and job descriptions, I suggest that at least for the task of deciding which experts to contact, there was a tendency for my participants to choose more reliable, mined sources rather than self-reported data, extending the value of signaling theory in differentiating between different sources of information available online.

### *Implications for practice*

This research was prompted by both academic and practical considerations. The already widespread proliferation of social computing technologies and the increasingly routine reliance on search systems indicate the growing importance of understanding how people use these systems and interpret the information they contain. The findings from this study should be valuable for practitioners wishing to understand how to best improve their people search interfaces and the data to include in such systems. Both interface design and organizational level recommendations are identified below.

Search engine companies are always interested in factors that influence clickthrough behavior. This is already evident from researchers looking at document search within this space (e.g. Clarke et al., 2007; Cutrell & Guan, 2007; Guan & Cutrell, 2007). The unique finding from the studies described in chapter 4 point towards displaying social connection snippets regarding mutual contacts and contacts that can be used as intermediaries. Indeed, the popular social networking website Facebook<sup>11</sup> has apparently already taken heed to this recommendation by displaying mutual contacts when searching for a person (Figure 19).

---

<sup>11</sup> <http://www.facebook.com>



Figure 19. Screenshot of Facebook interface showing 'mutual contacts'.

This dissertation also provides justification for aggregating and displaying social software participation data in expertise location systems. I have not come across many expertise locators that include or perform any systematic analysis on such data. Recent work has looked at how structural patterns within the social network of an online community can be used to identify ‘answer people’ (Welser, Gleave, Fisher, & Smith, 2007). An implication from this dissertation is that systematic analysis of participation in various forms of social software could be used to identify experts that are more likely to respond. This could be factored into search systems to create a ‘Page Rank’ for experts.

Finally, in this research I examined the likelihood of contacting an expert as a function of the rank order of the expert on a search results page, the social closeness of the expert to the participant and the degree of participation of the expert in visible social software tools such as blog posts, forums and social bookmarking. The actual response of the person contacted, ensuing interaction and its quality is a subject for future research. In part A of phase 2 in chapter 4, I looked at a single expertise search keyword to negate any confounding effects of the nature of expertise keyword. Some

keywords and experts in those areas, for instance those involved in esoteric aspects of compiler design, might not participate in social software. Future work will involve systematically varying the nature of the expertise keyword and determining its effect on whom a person decides to contact.



## REFERENCES

- Ackerman, M. S. (1994). *Augmenting the organizational memory: a field study of answer garden*. Paper presented at the CSCW '94: Proceedings of the 1994 ACM conference on Computer supported cooperative work, New York, NY.
- Ackerman, M. S. (1998). Augmenting organizational memory: a field study of answer garden. *ACM Transactions on Information Systems*, 16(3), 203-224.
- Ackerman, M. S., & Halverson, C. (2003). Sharing Expertise: The Next Step for Knowledge Management. In V. Wulf & M. H. Huysman (Eds.), *Social Capital and Information Technology*. . Cambridge, MA, USA: MIT Press.
- Ackerman, M. S., & Halverson, C. (2004). Sharing Expertise: The Next Step for Knowledge Management. In M. H. Huysman & V. Wulf (Eds.), *Social Capital and Information Technology* (pp. 273-299). Cambridge: MIT Press.
- Ackerman, M. S., & McDonald, D. W. (1996). *Answer Garden 2: merging organizational memory with collaborative help*. Paper presented at the 1996 ACM conference on Computer supported cooperative work, Boston, Massachusetts, United States.
- Ackerman, M. S., Starr, B., & Pazzani, M. (1997). *The Do-I-Care Agent: Effective Social Discovery and Filtering on the Web*. Paper presented at the RIAO'97: Computer-Assisted Information Searching on the Internet.
- Ackerman, M. S., Wulf, V., & Pipek, V. (2002). *Sharing Expertise: Beyond Knowledge Management*: MIT Press.
- Adamic, L. A. (1999). *The small world web*. Paper presented at the ECDL '99.

- Adamic, L. A., & Adar, E. (2005). How to search a social network. *Social Networks*, 27(3), 187-203.
- Adamic, L. A., Lukose, R. M., Puniyani, A. R., & Huberman, B. A. (2001). Search in power-law networks. *Physical Review E*, 64(4), 046135.
- Adar, E., & Adamic, L. A. (2005). *Tracking Information Epidemics in Blogspace*. Paper presented at the 2005 IEEE/WIC/ACM International Conference on Web Intelligence.
- Adler, P. S., & Kwon, S.-W. (2002). Social capital: Prospects for a new concept. *Academy of Management Review*, 27(1), 17-40.
- Akerlof, G. A. (1970). The market for lemons: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3), 488-500.
- Alexander, R. D. (1975). The Search for a General Theory of Behavior. *Behavioral Science*, 20, 77-100.
- Allen, T. J. (1977). *Managing the flow of technology*. Cambridge, MA: MIT Press.
- Anderson, J. R. (2000). *Cognitive psychology and its implications*. New York: Worth Publishers.
- Argote, L. (1999). *Organizational learning : creating, retaining, and transferring knowledge*. Boston: Kluwer Academic.
- Balabanovic, M. (1997). *An adaptive Web page recommendation service*. Paper presented at the The first international conference on Autonomous agents, Marina del Rey, California, United States.
- Balabanovic, M., & Shoham, Y. (1997). Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3), 66-72.

- Bates, M. (1989). The design of browsing and Berrypicking techniques for the online search interface. *Online Review*, 13(5), 407-442.
- Bauer, D. (2002). *Personal information geographies*. Paper presented at the CHI '02: CHI '02 extended abstracts on Human factors in computing systems, New York, NY.
- Baumgartner, F. R., & Leech, B. L. (1998). *Basic interests : the importance of groups in politics and in political science*. Princeton, N.J.: Princeton University Press.
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part 1. Background and theory. *Journal of Documentation*, 38(2), 61-71.
- Bergman, M., & Mark, G. (2002). *Technology choice as a first step in design: the interplay of procedural and sensemaking processes*. Paper presented at the DIS '02: Proceedings of the 4th conference on Designing interactive systems, New York, NY.
- Bernard, H. R., Killworth, P. D., & McCarty, C. (1982). Index: An informant-defined experiment in social structure. *Social Forces*, 61(1), 99-133.
- Bhavnani, S. K. (2005). Why is it difficult to find comprehensive information? Implications of information scatter for search and design: Research Articles. *Journal of the American Society for Information Science and Technology*, 56(9), 989-1003.
- Billman, D., & Bier, E. A. (2007). *Medical sensemaking with entity workspace*. Paper presented at the CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY.
- Bird, R. B., Smith, E. A., Alvard, M., & Chibnik, M. (2005). Signaling Theory,

- Strategic Interaction, and Symbolic Capital. *Current anthropology*, 46(2), 221-245.
- Bonacich, P. (1987). Power and Centrality: A Family of Measures. *American Journal of Sociology*, 92(5), 1170-1182.
- Borgatti, S. P., & Cross, R. (2003). A Relational View of Information Seeking and Learning in Social Networks. *Management Science*, 49(4), 432-445.
- Bos, N., Shami, N. S., Olson, J. S., Cheshin, A., & Nan, N. (2004). *In-group/out-group effects in distributed teams: an experimental simulation*. Paper presented at the ACM conference on Computer supported cooperative work 2004 Chicago, Illinois, USA.
- Bourdieu, P. (1977). *Outline of a theory of practice*. Cambridge [England]; New York: Cambridge University Press.
- Brass, D. J. (1992). Power in organizations a social network perspective. *Research in Politics and Society*, 4, 295-323.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7), 107-117.
- Brown, J. J., & Reingen, P. H. (1987). Social Ties and Word-of-Mouth Referral Behavior. *Journal of Consumer Research*, 14(3), 350-362.
- Burt, R. S. (2004). Structural holes and good ideas. *The American Journal of Sociology*, 110(2), 349-399.
- Carroll, J. M., & Rosson, M. B. (1992). Getting around the task-artifact cycle: how to make claims and design by scenario. *ACM Transactions on Information Systems*, 10(2), 181-212.

- Clark, H. H. (1996). *Using language*. New York: Cambridge University Press.
- Clarke, C. L. A., Agichtein, E., Dumais, S., & White, R. W. (2007). *The influence of caption features on clickthrough patterns in web search*. Paper presented at the SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY.
- Constant, D., Sproull, L., & Kiesler, S. (1996). The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice. *Organization Science*, 7(2), 119-135.
- Cronk, L. (2001). Anthropology and the Evolutionary Study of Human Behavior. *Research in Biopolitics*, 8, 1-30.
- Cronk, L. (2005). The application of animal signaling theory to human phenomena: some thoughts and clarifications. *Social Science Information*, 44(4), 603-620.
- Cross, R., & Borgatti, S. P. (2004). The ties that share: Relational characteristics that facilitate information seeking. In M. H. Huysman & V. Wulf (Eds.), *Social capital and information technology* (pp. 137-161). Boston, MA: MIT Press.
- Cross, R., & Sproull, L. (2004). More Than an Answer: Information Relationships for Actionable Knowledge. *Organization Science*, 15(4), 446-462.
- Cutrell, E., & Guan, Z. (2007). *What are you looking for?: an eye-tracking study of information usage in web search*. Paper presented at the CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY.
- Cutrell, E., & Guan, Z. (2007). *What are you looking for?: an eye-tracking study of*

- information usage in web search*. Paper presented at the CHI 2007.
- Dave, K., Wattenberg, M., & Muller, M. (2004). *Flash forums and forumReader: navigating a new kind of large-scale online discussion*. Paper presented at the CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work, New York, NY.
- DePaulo, B. M., & Fisher, J. D. (1980). The Costs of Asking for Help. *Basic & Applied Social Psychology*, 1(1), 23-35.
- Dervin, B. (1992). From the mind's eye of the user: The sense-making qualitative-quantitative methodology. In J. Glazier & R. Powell (Eds.), *Qualitative research in information management* (pp. 61-84). Englewood, CO: Libraries Unlimited.
- Dervin, B., & Dewdney, P. H. (1986). Neutral questioning: A new approach to the reference interview. *Reference Quarterly*, 25 (4), 506-513.
- Dodds, P. S., Muhamad, R., & Watts, D. J. (2003). An Experimental Study of Search in Global Social Networks. *Science*, 301, 827-829.
- Donath, J. (1999). Identity and deception in the virtual community. In M. A. Smith & P. Kollock (Eds.), *Communities in cyberspace* (pp. 29-59). London; New York: Routledge.
- Donath, J. (2007). Signals in Social Supernets. *Journal of Computer-Mediated Communication*, 13(1), 231-251.
- Donath, J. (In Press). *Signals, Truth & Design*. Cambridge, MA: MIT Press.
- Dorogovtsev, S. N., & Mendes, J. F. F. (2002). Evolution of networks. *Advances in Physics*, 51(4), 1079 - 1187.

- Ehrlich, K. (2003). Locating expertise: Design issues for an expertise locator system. In M. S. Ackerman, P. Volkmar & V. Wulf (Eds.), *Beyond Knowledge Management: Sharing Expertise*. Cambridge, MA: MIT Press.
- Ehrlich, K., Lin, C.-Y., & Griffiths-Fisher, V. (2007). *Searching for experts in the enterprise: combining text and social network analysis*. Paper presented at the GROUP '07: Proceedings of the 2007 international ACM conference on Supporting group work, New York, NY.
- Ehrlich, K., & Shami, N. S. (2008). *Searching for expertise*. Paper presented at the CHI '08: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY.
- Farrell, S., Lau, T., & Nusser, S. (2007). *Building Communities with People-Tags*. Paper presented at the INTERACT '07: Proceedings of the Eleventh IFIP TC13 International Conference on Human Computer Interaction, Berlin / Heidelberg.
- Farrell, S., Lau, T., Nusser, S., Wilcox, E., & Muller, M. (2007). *Socially augmenting employee profiles with people-tagging*. Paper presented at the UIST '07: Proceedings of the 20th annual ACM symposium on User interface software and technology, New York, NY.
- Fernald, A. (1992). Human Maternal Vocalizations to Infants as Biologically Relevant Signals: An Evolutionary Perspective. In J. Barkow, L. Cosmides & J. Tooby (Eds.), *The Adapted Mind* (pp. 391–428). Oxford: Oxford University Press.
- Fidel, R., & Green, M. (2004). The many faces of accessibility: engineers' perception of information sources. *Information Processing & Management*, 40(3), 563.
- Fiore, A. T., & Donath, J. (2005). *Homophily in online dating: when do you like*

- someone like yourself?* Paper presented at the CHI '05 extended abstracts on Human factors in computing systems.
- FitzGibbon, C. D., & Fanshawe, J. H. (1988). Stotting in Thomson's gazelles: an honest signal of condition. *Behavioral Ecology and Sociobiology*, 23(2).
- Fitzpatrick, G. (2003). Emergent Expertise Sharing in a New Community. In M. S. Ackerman, V. Pipek & V. Wulf (Eds.), *Sharing Expertise: Beyond Knowledge Management* (pp. 81-106). Cambridge, MA, USA: MIT Press.
- Foner, L. N. (1997). *Yenta: a multi-agent, referral-based matchmaking system*. Paper presented at the The first international conference on Autonomous agents, Marina del Rey, California, United States.
- Frank, R. H. (1988). *Passions within reason : the strategic role of the emotions*. New York: Norton.
- Frank, R. H. (2001). *Microeconomics and behavior*. New York: McGraw-Hill.
- Freeman, L. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, 1, 215-239.
- Furnas, G. W., & Russell, D. M. (2005). *Making sense of sensemaking*. Paper presented at the CHI '05: CHI '05 extended abstracts on Human factors in computing systems, New York, NY.
- Fussell, S. R., Kiesler, S., Setlock, L. D., & Scupelli, P. (2004). *Effects of instant messaging on the management of multiple project trajectories*. Paper presented at the SIGCHI conference on Human factors in computing systems 2004, Vienna, Austria.
- Gall, S. N.-L. (1985). Help-Seeking Behavior in Learning. *Review of Research in*



*Education 12*, 55-90.

Goffman, E. (1959). *The presentation of self in everyday life*. Garden City, N.Y.: Doubleday.

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using Collaborative Filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61-70.

Gosling, S. D., Gaddis, S., & Vazire, S. (2007). *Personality Impressions Based on Facebook Profiles*. Paper presented at the ICWSM '07: International conference on weblogs and social media.

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, 82(3), 379-398.

Gotz, D. (2007). *The ScratchPad: sensemaking support for the web*. Paper presented at the WWW '07: Proceedings of the 16th international conference on World Wide Web, New York, NY.

Grafen, A. (1990a). Biological signals as handicaps. *Journal of Theoretical Biology*, 144(4), 517-546.

Grafen, A. (1990b). Sexual selection unhandicapped by the fisher process. *Journal of Theoretical Biology*, 144(4), 473-516.

Granka, L., Joachims, T., & Gay, G. (2004). *Eye-tracking analysis of user behavior in WWW search*. Paper presented at the SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY.

- Granovetter, M. S. (1973). The Strength of Weak Ties. *The American Journal of Sociology*, 78(6), 1360-1380.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological review*, 102(1), 4-27.
- Guan, Z., & Cutrell, E. (2007). *An eye tracking study of the effect of target rank on web search*. Paper presented at the CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY.
- Guilford, T., & Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Animal Behaviour*, 42(1), 1-14.
- Hamilton, W. D., & Zuk, M. (1982). Heritable true fitness and bright birds: a role for parasites? *Science*, 218(4570), 384-387.
- Hancock, J. T., & Dunham, P. J. (2001). Impression Formation in Computer-Mediated Communication Revisited: An Analysis of the Breadth and Intensity of Impressions. *Communication Research*, 28(3), 325-347.
- Hancock, J. T., Toma, C., & Ellison, N. (2007). *The truth about lying in online dating profiles*. Paper presented at the CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY.
- Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly*, 44(1), 82-111.
- Harpending, H., Draper, P., & Rogers, A. (1987). Human Sociobiology. *Yearbook of Physical Anthropology*, 30, 127-150.
- Haythornthwaite, C. (2002). Strong, weak, and latent ties and the impact of new

media. *Information Society* 18(5), 385-401.

Herbsleb, J. D., Mockus, A., Finholt, T. A., & Grinter, R. E. (2000). *Distance, dependencies, and delay in a global collaboration*. Paper presented at the CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work, New York, NY.

Hertzum, M., & Pejtersen, A. M. (2000). The information-seeking practices of engineers: searching for documents as well as for people. *Information Processing & Management*, 36(5), 761-778.

Hill, W., & Terveen, L. G. (1996). *Using frequency-of-mention in public conversations for social filtering*. Paper presented at the CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work, Boston, MA, United States.

Hinde, R. A., & Barden, L. A. (1985). The evolution of the teddy bear. *Animal Behaviour*, 33(4), 1371-1373.

Hinds, P., Carley, K. M., Krackhardt, D., & Wholey, D. (2000). Choosing Work Group Members: Balancing Similarity, Competence, and Familiarity. *Organizational Behavior and Human Decision Processes*, 81(2), 226-251.

Hinds, P., & Pfeffer, J. (2003). Why Organizations Don't "Know What They Know": Cognitive and Motivational Factors Affecting the Transfer of Expertise. In M. S. Ackerman, V. Pipek & V. Wulf (Eds.), *Sharing expertise: Beyond knowledge management*. Cambridge, MA: MIT Press.

Hollingshead, A., Fulk, J., & Monge, P. (2002). Applying the theories of transactive memory and public goods to intranets and knowledge management in

- organizations. In P. Hinds & S. Kiesler (Eds.), *Distributed work: New research on working across distance using technology* (pp. 335-355). Cambridge, MA: MIT Press.
- Huh, J., Jones, L., Erickson, T., Kellogg, W. A., Bellamy, R. K. E., & Thomas, J. C. (2007). *BlogCentral: the role of internal blogs at work*. Paper presented at the CHI '07: CHI '07 extended abstracts on Human factors in computing systems, New York, NY.
- Ibarra, H. (1992). Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm. *Administrative Science Quarterly*, 37(3), 422-447.
- Jehn, K. A., & Shah, P. P. (1997). Interpersonal relationships and task performance: An examination of mediation processes in friendship and acquaintance groups. *Journal of Personality and Social Psychology*, 72(4), 775-790.
- Joachims, T., Granka, L., Pan, B., & Gay, G. (2005). *Accurately interpreting clickthrough data as implicit feedback*. Paper presented at the SIGIR 2005.
- Joachims, T., Granka, L., Pan, B., Hembrooke, H., Radlinski, F., & Gay, G. (2007). Evaluating the accuracy of implicit feedback from clicks and query reformulations in Web search. *ACM Transactions on Information Systems*, 25(2).
- Kautz, H., Selman, B., & Shah, M. (1997a). Referral Web: combining social networks and collaborative filtering. *Communications of the ACM*, 40(3), 63-65.
- Kautz, H., Selman, B., & Shah, M. (1997b). The hidden web. *AI Magazine*, 18, 27-36.
- Killworth, P., & Bernard, H. R. (1978). Reverse small world experiment. *Social*

- Networks*, 1, 159-192.
- Klein, B., & Leffler, K. B. (1981). The Role of Market Forces in Assuring Contractual Performance. *Journal of Political Economy*, 89(4), 615.
- Klein, G., Moon, B., & Hoffman, R. R. (2006a). Making Sense of Sensemaking 1: Alternative Perspectives. *Intelligent Systems*, 21(4), 70-73.
- Klein, G., Moon, B., & Hoffman, R. R. (2006b). Making Sense of Sensemaking 2: A Macrocognitive Model. *Intelligent Systems*, 21(5), 88-92.
- Kleinberg, J. M. (2000). Navigation in a small world. *Nature*, 406(6798), 845.
- Kollman, K. (1998). *Outside lobbying : public opinion and interest group strategies*. Princeton, N.J.: Princeton University Press.
- Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). GroupLens: applying collaborative filtering to Usenet news. *Communications of the ACM*, 40(3), 77-87.
- Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind reading and manipulation. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology : an evolutionary approach* (2nd ed.). Sunderland, MA: Sinauer Associates.
- Lampe, C., Ellison, N., & Steinfield, C. (2006). *A face(book) in the crowd: social Searching vs. social browsing*. Paper presented at the CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, New York, NY.
- Lave, J., & Wenger, E. (1991). *Situated learning : legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Lee, F. (2002). The Social Costs of Seeking Help. *Journal of Applied Behavioral*

*Science*, 38(1), 17-35.

Leonard-Barton, D., & Sinha, D. K. (1993). Developer-User Interaction and User Satisfaction in Internal Technology Transfer. *The Academy of Management Journal*, 36(5), 1125-1139.

Lin, C.-Y., Ehrlich, K., Griffiths-Fisher, V., & Desforges, C. (2008). SmallBlue: People Mining for Expertise Search. *IEEE Multimedia Magazine*, 15(1), 78-84.

Lipshitz, R., Klein, G. A., Orasanu, J., & Salas, E. (2001). A welcome dialogue--and the need to continue. *Journal of Behavioral Decision Making*, 14(5), 385-389.

Littlepage, G. E., Schmidt, G. W., Whisler, E. W., & Frost, A. G. (1995). An input-process-output analysis of influence and performance in problem-solving groups. *Journal of Personality and Social Psychology*, 69(5), 877-889.

Louis, M. (1980). Surprise and Sensemaking: What Newcomers Experience in Entering Unfamiliar Organizational Settings. *Administrative Science Quarterly*, 25, 226-251.

Majchrzak, A., Wagner, C., & Yates, D. (2006). *Corporate wiki users: results of a survey*. Paper presented at the WikiSym '06: Proceedings of the 2006 international symposium on Wikis, New York, NY.

Malone, T. W., Grant, K. R., Lai, K.-Y., Rao, R., & Rosenblitt, D. A. (1989). The information lens: An intelligent system for information sharing and coordination. In M. H. Olson (Ed.), *Technical support for work group collaboration* (pp. 65-88). Hillsdale, NJ: Lawrence Erlbaum.

Maltz, D., & Ehrlich, K. (1995). *Pointing the way: active collaborative filtering*. Paper

presented at the SIGCHI conference on Human factors in computing systems  
1995, Denver, Colorado, United States.

Manning, C. D., & Schutze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, Mass.: MIT Press.

Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). *HT06, tagging paper, taxonomy, Flickr, academic article, to read*. Paper presented at the  
HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext  
and hypermedia, New York, NY.

Mauss, M. (1954). *The Gift: The Form and Reason for Exchange in Archaic Societies*.  
New York: Routledge.

Maynard Smith, J. (1991). Honest signalling: the Philip Sidney game. *Animal Behaviour*, 42(6), 1034-1035.

Maynard Smith, J., & Harper, D. (2003). *Animal signals*. New York: Oxford  
University Press.

Maynard Smith, J., & Harper, D. G. C. (1995). Animal Signals: Models and  
Terminology. *Journal of Theoretical Biology*, 177(3), 305-311.

McAllister, D. J. (1995). Affect- and Cognition-Based Trust as Foundations for  
Interpersonal Cooperation in Organizations. *The Academy of Management Journal*, 38(1), 24-59.

McArthur, R., & Bruza, P. (2003). *Discovery of Implicit and Explicit Connections between People Using Email Utterance*. Paper presented at the Conference on  
Computer Supported Cooperative Work 2003, Helsinki, Finland.

McDonald, D. W. (1999). *Recommending expertise in an organizational setting*. Paper

presented at the CHI '99 extended abstracts on Human factors in computing systems, Pittsburgh, Pennsylvania.

McDonald, D. W., & Ackerman, M. S. (1998). *Just talk to me: A field study of expertise location*. Paper presented at the CSCW '98: Proceedings of the 1998 ACM conference on Computer supported cooperative work, New York, NY.

McDonald, D. W., & Ackerman, M. S. (2000). *Expertise recommender: a flexible recommendation system and architecture*. Paper presented at the CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work, New York, NY.

McGinniss, J. (1983). *Fatal vision*. New York: Putnam.

Menczer, F. (2002). Growing and navigating the small world Web by local content. *Proceedings of the National Academy of Sciences of the United States of America*, 14014-14019.

Mika, P. (2005). Flink: Semantic Web technology for the extraction and analysis of social networks. *Journal of Web Semantics*, 3(2-3), 211-223.

Mika, P., Elfring, T., & Groenewegen, P. (2006). Application of semantic technology for social network analysis in the sciences. *Scientometrics*, 68(1), 3-27.

Millen, D. R., Feinberg, J., & Kerr, B. (2006). *Dogear: Social bookmarking in the enterprise*. Paper presented at the CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems, New York, NY.

Moreland, R., Argote, L., & Krishnan, R. (1996). Socially shared cognition and work: Transactive memory and group performance. In J. L. Nye & A. M. Brower (Eds.), *What's social about social cognition? research on socially shared*



- cognition in small groups* (pp. 57). Thousand Oaks, Calif.: Sage Publications.
- Morrison, E. W., & Vancouver, J. B. (2000). Within-person analysis of information seeking: the effects of perceived costs and benefits. *Journal of Management*, 26(1), 119-137.
- Muck, T., & West, R. (2004). *Dreamweaver MX 2004 the complete reference*. Berkeley, Calif.; London: Osborne ; McGraw-Hill.
- Nardi, B. A., Whittaker, S., & Schwarz, H. (2002). It's not what you know, it's who you know: Work in the information age [Electronic Version]. *First Monday*, from [www.firstmonday.org](http://www.firstmonday.org)
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2), 167-256.
- Newman, M. E. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, 27(1), 39-54.
- Newman, M. E. J., & Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E*, 68(3), 1063-1071.
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company : how Japanese companies create the dynamics of innovation*. New York: Oxford University Press.
- Norton, E. C., Bieler, G. S., Ennett, S. T., & Zarkin, G. A. (1996). Analysis of prevention program effectiveness with clustered data using generalized estimating equations. *Journal of Consulting and Clinical Psychology*, 64(5), 919-926.
- Olson, G. M., & Olson, J. S. (2000). Distance Matters. *Human-Computer Interaction*,

15(2), 139-178.

Oreg, S., & Nov, O. (In Press). Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Computers in Human Behavior, In Press, Corrected Proof*.

Orr, J. E. (1996). *Talking about machines : an ethnography of a modern job*. Ithaca, NY, USA: Cornell University Press.

Paepcke, A. (1996). Information needs in technical work settings and their implications for the design of computer tools. *Computer Supported Cooperative Work, 5*(1), 63-92.

Palazzolo, E. T. (2005). Organizing for Information Retrieval in Transactive Memory Systems. *Communication Research, 32*(6), 726-761.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google We Trust: Users' Decisions on Rank, Position, and Relevance. *Journal of Computer-Mediated Communication, 12*(3), 801-823.

Penuel, B., & Cohen, A. (2003). Coming to the crossroads of knowledge, learning, and technology: Integrating knowledge management and workplace learning. In M. S. Ackerman, V. Pipek & V. Wulf (Eds.), *Sharing expertise: Beyond knowledge management* (pp. 57-76). Cambridge, MA: MIT Press.

Pipek, V., & Wulf, V. (2003). *Pruning the answer garden: Knowledge sharing in maintenance engineering*. Paper presented at the Proceedings of the eighth conference on European Conference on Computer Supported Cooperative Work.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological review, 106*(4),

643-675.

- Qu, Y. (2003). *A sensemaking-supporting information gathering system*. Paper presented at the CHI '03: CHI '03 extended abstracts on Human factors in computing systems, New York, NY.
- Qu, Y. (2006). *Supporting representation construction in sensemaking*. Unpublished PhD Dissertation, University of Michigan Ann Arbor.
- Qu, Y., & Furnas, G. W. (2005). *Sources of structure in sensemaking*. Paper presented at the CHI '05: CHI '05 extended abstracts on Human factors in computing systems, New York, NY.
- Reichling, T., Schubert, K., & Wulf, V. (2005). *Matching human actors based on their texts: design and evaluation of an instance of the ExpertFinding framework*. Paper presented at the 2005 international ACM SIGGROUP conference on Supporting group work, Sanibel Island, Florida, USA.
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). *GroupLens: an open architecture for collaborative filtering of netnews*. Paper presented at the Proceedings of the 1994 ACM conference on Computer supported cooperative work.
- Riegelsberger, J., Counts, S., Farnham, S. D., & Philips, B. C. (2006). *Sounds good to me: effects of photo and voice profiles on gaming partner choice*. Paper presented at the Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, Banff, Alberta, Canada.
- Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour*, 58(5), 921-931.

- Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K. (1993). *The cost structure of sensemaking*. Paper presented at the CHI '93: Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems, New York, NY.
- Sarwar, B. M., Konstan, J. A., Borchers, A., Herlocker, J. L., Miller, B. N., & Riedl, J. (1998). *Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system*. Paper presented at the Proceedings of the 1998 ACM conference on Computer supported cooperative work.
- Schelling, T. C. (1960). *The strategy of conflict*. Cambridge: Harvard University Press.
- Shami, N. S., Bos, N., Wright, Z., Hoch, S., Kuan, K. Y., Olson, J. S., et al. (2004). *An experimental simulation of multi-site software development*. Paper presented at the Proceedings of the 2004 conference of the Centre for Advanced Studies on Collaborative research.
- Shami, N. S., Ehrlich, K., & Millen, D. R. (2008). *Pick me! Link selection in expertise search results*. Paper presented at the CHI '08: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY.
- Shami, N. S., Yuan, Y. C., Cosley, D., Xia, L., & Gay, G. (2007). *That's what friends are for: facilitating 'who knows what' across group boundaries*. Paper presented at the GROUP '07: Proceedings of the 2007 international ACM conference on Supporting group work, New York, NY.
- Shapiro, E. G. (1980). Is Seeking Help from a Friend Like Seeking Help from a Stranger? *Social Psychology Quarterly* 43(2), 259-263.

- Sharma, N. (2007). *Sensemaking handoff: theory and recommendations*. Paper presented at the CHI '07: CHI '07 extended abstracts on Human factors in computing systems, New York, NY.
- Simon, H. A. (1972). *The sciences of the artificial*. Cambridge, MA.: M. I. T. Press.
- Spence, M. (1973). Job Market Signaling. *Quarterly Journal of Economics*, 87(3), 355-374.
- Starbuck, W. H., & Milliken, F. J. (1988). Executives' Perceptual Filters: What They Notice and How They Make Sense. In D. C. Hambrick (Ed.), *The executive effect: Concepts and methods for studying top managers* (pp. 35-65). Greenwich, CT: JAI.
- Stecher, K. B., & Counts, S. (2008). *Spontaneous inference of personality traits from online profiles*. Paper presented at the International Conference on Weblogs and Social Media
- Streeter, L. A., & Lochbaum, K. E. (1988a). *An expert/expert-locating system based on automatic representation of semantic structure*. Paper presented at the Fourth conference on artificial intelligence applications, San Diego, CA.
- Streeter, L. A., & Lochbaum, K. E. (1988b). *Who knows: A system based on automatic representation of semantic structure*. Paper presented at the RIAO '88, Cambridge, MA.
- Taylor, R. S. (1962). The process of asking questions. *American Documentation*, 13(4), 391-396.
- Terveen, L. G., Hill, W., Amento, B., McDonald, D., & Creter, J. (1997). PHOAKS: a system for sharing recommendations. *Communications of the ACM*, 40(3), 59-

62.

- Terveen, L. G., & McDonald, D. (2005). Social matching: A framework and research agenda. *ACM Transactions on Computer Human Interaction (TOCHI)*, *12*(3), 401-434.
- Terveen, L. G., Selfridge, P. G., & Long, M. D. (1995). Living Design Memory: Framework, Implementation, Lessons Learned. *Human-Computer Interaction*, *10*(1), 1-38.
- Thom-Santelli, J., Muller, M. J., & Millen, D. R. (2008). *Social tagging roles: publishers, evangelists, leaders*. Paper presented at the Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, Florence, Italy.
- Thomas, J. B., Clark, S. M., & Gioia, D. A. (1993). Strategic Sensemaking and Organizational Performance: Linkages among Scanning, Interpretation, Action, and Outcomes. *Academy of Management Journal*, *36*(2), 239-270.
- Travers, J., & Milgram, S. (1969). Experimental Study Of Small World Problem. *Sociometry*, *32*(4), 425-443.
- Trivers, R. L. (1974). Parent-Offspring Conflict. *American Zoologist*, *14*(1), 249-264.
- Uleman, J. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 141-160). New York: Guilford Press.
- Vazire, S., & Gosling, S. D. (2004). e-Perceptions: Personality Impressions Based on Personal Websites. *Journal of Personality and Social Psychology*, *87*(1), 123-132.

- Veblen, T. (1965). *The theory of the leisure class. 1899*. New York: A.M. Kelley bookseller.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Watts, D. J., Dodds, P. S., & Newman, M. E. J. (2002). Identity and Search in Social Networks. *Science*, 296(5571), 1302.
- Wegner, D. M. (1987). Transactive memory: A contemporary analysis of group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185). New York: Springer-Verlag.
- Wegner, D. M. (1995). A computer network model of human transactive memory. *13*(3), 319.
- Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks: Sage Publications.
- Weisz, J. D., Erickson, T., & Kellogg, W. A. (2006). *Synchronous broadcast messaging: the use of ICT*. Paper presented at the CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems, New York, NY.
- Welser, H., Gleave, E., Fisher, D., & Smith, M. A. (2007). Visualizing the Signatures of Social Roles in Online Discussion Groups. *Journal of Social Structure*, 8(2), <http://www.cmu.edu/joss/content/articles/volume8/Welser>.
- Westley, F. R. (1990). Middle Managers and Strategy: Microdynamics of Inclusion. *Strategic Management Journal*, 1(1), 337-351.
- Whittaker, S. (2008). Making Sense of Sense Making. In T. Erickson & D. McDonald (Eds.), *HCI Remixed* (pp. 173-178). Cambridge, MA, USA: MIT Press.

- Wiseman, J. P. (1986). Friendship: Bonds and Binds in a Voluntary Relationship. *Journal of Social and Personal Relationships*, 3(2), 191-211.
- Woudstra, L., & van den Hooff, B. (2008). Inside the source selection process: Selection criteria for human information sources. *Information Processing & Management*, 44(3), 1267-1278.
- Yimam-Seid, D., & Kobsa, A. (2003). Expert-Finding Systems for Organizations: Problem and Domain Analysis and the DEMOIR Approach. *Journal of Organizational Computing and Electronic Commerce*, 13(1), 1.
- Yu, B., & Singh, M. P. (2003). *Searching social networks*. Paper presented at the Second international joint conference on Autonomous agents and multiagent systems, Melbourne, Australia.
- Yuan, Y. C., Fulk, J., & Monge, P. (2007). Access to Information in Connective and Communal Transactive Memory Systems. *Communication Research*, 34(2), 131-155.
- Yuan, Y. C., Fulk, J., Shumate, M., Monge, P. R., Bryant, J. A., & Matsaganis, M. (2005). Individual Participation in Organizational Information Commons: The Impact of Team Level Social Influence and Technology-Specific Competence. *Human Communication Research*, 31(2), 212-240.
- Zahavi, A. (1975). Mate selection - A selection for a handicap. *Journal of Theoretical Biology*, 53(1), 205-214.
- Zahavi, A., & Zahavi, A. (1997). *The handicap principle a missing piece of Darwin's puzzle*. New York: Oxford University Press.
- Zhang, J., & Ackerman, M. S. (2005). *Searching for expertise in social networks: a*



*simulation of potential strategies*. Paper presented at the 2005 international ACM SIGGROUP conference on Supporting group work, Sanibel Island, Florida, USA.

Zuk, M., Ligon, J. D., & Thornhill, R. (1992). Effects of experimental manipulation of male secondary sex characters on female mate preference in red jungle fowl. *Animal Behaviour*, 44(6), 999-1006.