Contextual Authority Tagging:
Expertise Location via Social Labeling

Terrell G. Russell

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Information and Library Science.

Chapel Hill
2011

Approved by:

Deborah Barreau, Advisor

Gary Marchionini, Advisor

Barbara M. Wildemuth, Committee Member

Sri Kalyanaraman, Committee Member

Phillip J. Windley, Committee Member

Abstract

Terrell G. Russell: Contextual Authority Tagging:
Expertise Location via Social Labeling.
(Under the direction of Deborah Barreau and Gary Marchionini.)

This study investigates the possibility of a group of people making explicit their tacit knowledge about one another's areas of expertise. Through a design consisting of a modified Delphi Study, group members are asked to label both their own and each others' areas of expertise over the course of five rounds. Statistical analysis and qualitative evaluation of 10 participating organizations suggest they were successful and that, with simple keywords, group members can convey the salient areas of expertise of their colleagues to a degree that is deemed "similar" and of "high quality" by both third parties and those being evaluated. More work needs to be done to make this information directly actionable, but the foundational aspects have been identified.

In a world with a democratization of voices from all around and increasing demands on our time and attention, this study suggests that simple, aggregated third-party expertise evaluations can augment our ongoing struggle for quality information source selection. These evaluations can serve as loose credentials when more expensive or heavyweight reputation cues may not be viable.

Acknowledgements

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Today, the Internet has democratized speech at every level. It has made free and open speech more available to everyone but it has not provided us with the requisite filters to disambiguate the signal from all the new noise. For democratic purposes, it is important that everyone have a voice (and an equal vote), but for most other purposes, it is not necessary. For most purposes, it is most helpful to hear the opinions of those who know what they are talking about and who have the most to offer the conversation. Reliably knowing who the experts are would be the first step of a larger goal to filter the signal from the noise in our Internet-empowered world where everyone can have a bullhorn.

The fundamental issue of expertise location has been faced at a smaller-than-web scale within individual organizations. Knowing what an organization knows about, and who carries that knowledge, is a valuable asset and has been a primary focus of knowledge management for many years. In large part,

knowing who knows what has come from two places – the individuals who have self-reported their own expertise and from algorithmic derivation from the produced documents and paper trail of doing business.

I think that a valuable third source is being overlooked. I think that people, other than the individual, have interesting insight and knowledge about what the individual knows. I think that their collective human opinion can serve as a reliable indicator of knowledge as well and should be included.

This dissertation research evaluates the ability of a group to know and express what an individual knows.

1.2   Overview

Humans can only sense and process so much. Because of this physical limitation, we have sought shortcuts in order to help us sense "more" (Downs, 1957) and to make up for our limited ability to have encyclopedic knowledge of the situations around us (Lupia, 1994). The use of many of these shortcuts is dependent on other people – those who are around us now, but also those who have come before, and those who are far away. Our dependence on others is inefficient in that we do not always know whom to ask or approach for help. Sometimes we waste valuable time and energy looking for the right source of information. We may be able to reduce this waste with some thoughtful sharing and collective reflection. We could benefit greatly by discovering the latent, undocumented knowledge of those around us and bringing it to the surface. We should be able to tap the implicit by making it more explicit

(Nonaka, 1991).

This research is an investigation into how a group of people can come to know what it is that its members know. Through simple keyword tagging and cognitive reflection on those tags over time, an individual and a group of his or her peers may approach a common ground or "shared understanding" around the topic of his or her areas of expertise. Better senses of self-awareness, other-awareness, and downstream decision-making may come about because of this information being collected and shared. This research is primarily focused on tagging data around humans whose granted cognitive authority (Wilson, 1983) to one another changes over time.

1.3   Problem Statement

Knowledge of our surroundings, from an empiricist perspective, comes from our five senses. The things we see and hear, the things we smell and touch and taste, they are all just constrained representations of our environment. We strive to make as much sense of the world as possible, but we are limited by our physical location, our position in time, access to information resources, and by the processing power of our brains (Dervin, 1983). Cognitive load theory (Sweller, Merrienboer, & Paas, 1998) tells us that we can only handle so much data coming in at a time.

Because of this constraint, we seek shortcuts, or second-hand information, in order to "see" more, to see beyond what is readily apparent. We seek shortcuts in order to "know" more than what our senses can sense. I think

these pieces of second-hand information can be of two distinct types, either basic pieces of simple information, or information that resembles an executive summary. Second-hand information can come from others in the form of basic facts such as "it's raining outside" or "it's raining at the beach" – both of which are simple facts but relayed to us by another, rather than collected or sensed on our own. Second-hand information can also come in the form of more summarized or processed information like "our economy is in a recession." This second type of new information could have been determined by one person or synthesized by many, but it also comes from sources outside of ourselves and is then relayed to us. Most of our information about the world is actually acquired this way – as second hand knowledge (Wilson, 1983). We experience firsthand very little of what we come to "know."

We depend on processing and sense-making done by others, in a different place, in a different time, to help us make sense of our world (sometimes to a polarizing degree (Gilovich, 1987)). This outsourcing of sense-making is fueled by necessity. We do not have the time or energy to collect, process, synthesize, and employ all our own data in a modern world. There is a division of labor and with it a division of knowledge and expertise (who was the last person to know "everything"?). To function in a (modern?) society, we depend on others, both past and present, for help when fulfilling our information needs.

With this dependence on others, both in person and via the documents and records others create, we must also be wary. We must keep a vigilant eye towards the legitimacy of the information being passed along. We must evaluate, critically, the source and the provenance of second-hand information.

Savolainen writes that when evaluating others and what we think they know, "overall, cognitive authority was characterized as having six facets; trustworthiness, reliability, scholarliness, credibility, 'officialness' and authoritativeness; of these, trustworthiness was perceived as the primary facet" (Savolainen, 2007, 3).

Even with the successful vetting and application of second-hand information, or shortcuts, from others, we never have perfect information. We may collect more information and we may collect better information, but it is never all the information we need to make perfect decisions. We satisfice; we satisfy with what is sufficient (Simon, 1957). We use what information we have to make decisions that we deem to be good enough at the time. We often seek out more information before making a decision but we have, what Simon called, "bounded rationality." We have imperfect information, limited attention and money, limited processing power and limited time, but we still need to make decisions.

Choo's Decision Behavior Model shows us that contextualized decision making happens within organizations based on cognitive limits, information quality and availability, and the values of the organization (Choo, 1996, 332). These inputs are handled with bounded rationality and within the confines of performance concerns, and whether the decision is good enough, among other simplifications. This decision making behavior is both rationally expected and observed.

Even knowing we will never have perfect information when working in these limited environments, we can arguably make better decisions if we can

improve or increase the amount of information on hand when making decisions. Having more good information reduces uncertainty about the environment surrounding a decision, but it does not necessarily reduce equivocality. To reduce equivocality, or ambiguity, of the information we have on hand, we need sensemaking and a perspective that comes from "retrospective interpretations" of earlier data and decisions (Choo, 1996, 334). We need to have *seen this before* and *know what it means*. What we need to make good decisions, in addition to good information, is called expertise.

There is a vast amount of latent, untapped information in the environment around us. Some of it is in the built world, some of it is in the natural world (too big, too small, hidden in non-visible wavelengths, etc.), and some of it is in the heads of those around us. Cross and Sproull (2004) noted that 85% of managers immediately mentioned specific people when asked "to describe sources of information important to successful completion of their project". They went on to write:

> As one manager said, "I mean the whole game is just being the person that can get the client what they need with [the Firm's] resources behind you. This almost always seems to mean knowing who knows what and figuring out a way to bring them to your client's issue" (R6). Very few of the named people were simply organizationally designated "experts"; most were described as partners in information relationships.

If we are informed by the right people before making decisions, and they help us decide what we are looking for (Belkin, Oddy, & Brooks, 1982), then we may improve our knowledge and understanding of a situation or problem at the time when we need to decide. **Knowing from whom we should get**

our information, when we are not sure of what we need, is a hard problem.

Expertise location, for this reason, has been a focus of the knowledge management field for many years. Knowledge management has also focused on the process of organizational learning and dissemination of that learning within the organization. In many cases, this has been done through the tracking of created documents and other knowledge artifacts (Martin, 2008).

An additional approach should consist of uncovering that which has not *yet* been recorded – that information which is in the heads of a group's membership. We should be equipped to hold up a mirror to help reflect an organization's insights and expertise back on itself. We need to help uncover the dark corners where we are not sure about the expertise in the room. With a regimen of self-reflection, iterated over time, I hope this problem can be made less hard. I think we can discover *whom to ask* for the relatively low cost of a little sustained individual effort and some focused record-keeping in the distributed network.

1.4   Significance

When we are seeking answers to questions or trying to increase our knowledge in a certain domain, we seek sources of information that are credentialed and tested. We ask those who have come before us and who have learned from their own experiences – either through doing or through their own process of seeking and discovery. The sources we come to trust should have a history of

providing good information in that domain in the past. We also come to expect them to continue to provide good information into the future. They should be known by others as keepers of good information and sound provenance. Our highly concentrated word for this set of qualities is reputation.

Those who have a good reputation perhaps spent many years developing their stature or physical skills in a field or domain. From the world of archival studies[1,2], we know that physical and electronic sources of information should have a clear chain of custody and line of provenance as the document of record. If people are to be trusted as sources, as experts, we should be able to see the clear chain of custody and provenance of those who defer to these experts. Identifying these trusted human sources and the provenance to go with them is the thrust of this research.

Knowledge management has been about having the organization know what its members know. If this is synthesized a bit, we may talk of what the members *know about*. If we can reliably assume that a group can know what a person knows about, we can potentially do some very interesting things. We may be able to render moot the concerns we have today with individuals lying to increase their stature. If the group can reliably increase the social friction necessary to gain unmerited influence, we could safely ignore the opinions of those who have not convinced quite a few of his peers that he knows what he is talking about. In a world where we do care about credentials, until

---

[1]Society of American Archivists' *Archives, Personal Papers, and Manuscripts*

[2]Canadian Council of Archives' *Rules for Archival Description*

one has at least the loose credential of a few peers who vouch for his credibility, one's potential for abusing that credibility is severely limited. Of course, existing credentials, even more formal credentials (diplomas, certifications, licensing, etc.), already allow this kind of credibility abuse. The addition of a loose socially awarded credential to the existing landscape would not affect the potential for abuse of those existing formal credentials. One would assume they would continue to convey more credibility than that provided by social labeling alone.

If a group can know what areas of expertise a person has, it may be able to better distribute articles for peer review to those who can best ascertain the quality of a pending publication. Important questions that arise could be distributed more reliably to those who could provide an informed opinion. Reporters in remote locations may have been better able to determine who had actually been on the ground during the 2010 presidential elections in Iran and who has recently created a Twitter account only to influence the placement of news articles during the next news cycle.

In a more formalized decision making process, voting systems could have weighted votes. If the matter at hand should not be decided strictly democratically (e.g., one person, one vote), the relative weight of the votes could be set to match the relative weight of a voter's apparent relative expertise on the matter. This could mirror the practice of corporate elections based on shareholder totals. Those who know, instead of those who own, would be rewarded with influence. Perhaps just as interestingly, those who do not know could be ignored at vote-tallying time. Internet-scale applications are often fraught

with noisy comments and hostility. These could be programatically tuned out or weighted less if it was deemed useful or helpful to do so. And this could be done site-wide or customized for each viewer based on personal taste. It is also important to note that this type of filtering would be done post-hoc. It would not affect who could initially vote, comment, or otherwise share their opinion. It would only affect how the display of the event would be rendered later. The original "democratic" vote totals would still be tallied and available.

But all of these scenarios depend on the assumption that a group's opinion about a member's areas of expertise can be trusted as "correct" – as good enough. The group's visible, shared opinion should allow the members of the group to make better, more informed decisions with less effort in less time.

I want to provide a robust means for allowing a group to assess and believe in their collective opinion about an individual's areas of expertise. They would be able to transparently evaluate how they grant cognitive authority to an individual and continually reflect on it. It would become a market indicator of what people know – one that fits into a larger, existing ecosystem.

This social reflecting lens should provide a form of **loose credentialing** and help to bring the implicit to the surface and make it explicit. When they choose to provide it, the trusted, focused, tacit knowledge in the heads of those we know could be available to all of us.

> "The grand challenge is to boost the collective IQ of organizations and of society" - Doug Engelbart regarding the Bootstrap Principle, a human-machine system for harvesting collected knowledge and evolving the technology for collective learning (Engelbart, 2004)

We still have far to go before the online and offline worlds truly merge. Eventually, we will enjoy a global transparent layer of data that is collectively curated and managed, but until that time, we continue to interact with other humans face-to-face much more often and in much more significant capacities. Lowenstein says that people trust their offline counterparts more than online social media (Lowenstein, 2009). However, research in computer-mediated communication (CMC) says we react to machines as people, at least subconsciously (Reeves & Nass, 1996), but we still have deference towards "real people" when we take the time to think through the communication event more carefully. When interacting with others via mediated channels, we usually do not focus on the medium itself and therefore we confer trust more than when the medium is explicitly obvious to us. As a medium continues to become more transparent and easy and common, it will become more trusted.

1.5 Related Work

1.5.1 Expertise Location

Organizational Memory (OM) is a key component of Knowledge Management (KM). Abecker, Bernardi, Hinkelmann, Kühn, and Sintek (1997, 1) write "that an OM [system] has to be more than an information system but must help to transform information into action." One part of OM is Expertise Location and Management (ELM), or the tracking of know-how within an organization (Lamont, 2003). As keeping track of employees' knowledge is generally a very expensive undertaking for any size organization, a cheaper, more efficient

technique for uncovering, managing, and disseminating this type of information would be a key contribution.

KM exercises involving human time and effort are naturally expensive for the firm. As such, incentivizing participation is one of the greatest hurdles to the implementation of a KM system (Ehrlich, 2003). Engaging with professional communities of practice (Lave & Wenger, 1991; Duguid, 2005), physical workspace reconfiguration, and encouraging water-cooler discussions can each improve the sharing and awareness of expertise among professionals. Even so, Ling, Sandhu, and Jain (2009, 135) suggest that the single best type of incentives for knowledge sharing activities remain top-down such as "rewards and performance appraisal." Callahan agrees and suggests that managers must be involved, resources (time and money) must be given to the task, and overt (already known) content must be used to seed any initial system that hopes to elicit tacit content (Callahan, 2006b).

Stein (1995) provides a standard set of stages for the understanding of organizational memory - knowledge acquisition, retention, maintenance, and retrieval. This is similar, but not identical, to Dieng's model for corporate memory management - detection of needs, knowledge construction, distribution, use, evaluation, and evolution (Dieng, Corby, Giboin, & Ribiére, 1999). Each suggests a timelined progression but differ in that Stein's stages feel more institutionalized and less a collaborative effort. Dieng's use, evaluation, and evolution incorporate the dynamic nature and multi-person aspects of a distributed know-how.

Dieng et al. (1999, 578) write:

However, the goal of a corporate memory building is different from the goal of an expert system: instead of aiming at an automatic solution for a task (with automatic reasoning capabilities), a corporate memory rather needs to be an assistant to the user, supplying him/her with relevant corporate information but leaving him/her the responsibility of a contextual interpretation and evaluation of this information (Kühn & Abecker, 1997). Kühn and Abecker (1997) notices that 'in contrast to expert systems, the goal of a corporate memory is not the support of a particular task, but the better exploitation of the essential corporate resource: knowledge' and cites some knowledge-based corporate memories (e.g., KONUS system aimed at support to crankshaft design).

Existing tools around Expertise Location and Management involve, almost entirely, self-description or existing-document data-mining (Lamont, 2003; Fitzpatrick, 2001; Becks, Reichling, & Wulf, 2004; Balog, Azzopardi, & Rijke, 2009). Traditional tf–idf[3] and bag-of-words analysis on these document stores can uncover a vast amount, but I think these techniques are missing out on what is in the heads of those who work with the person of interest. This is an important enough distinction to be made in a controlled environment, where the identities of the people involved are fairly well known and stable. However, trusting self-description in an unstructured, internet-wide environment without corporate identity management software seems ripe for abuse. The individual in question could easily be misrepresenting him or herself with malicious intent. Convincing many others of a lie or getting others to lie in a consistent manner regarding one's areas of expertise is much harder than deciding to lie on one's own behalf.

_____

[3]Term Frequency–Inverse Document Frequency

A tool for assisting in Expertise Location should meet the following requirements as set forth by Abecker et al. (1997):

- gather information from multiple sources

- integrate with existing infrastructure and practices

- require little overhead in time/attention and provide benefits quickly

- actively present relevant information

- must stay up-to-date

Contextual Authority Tagging would handle the first, third, and fifth natively. Integration and presentation would both depend on implementation details. Ehrlich goes on to say that these systems must be fast, easy to use, engender trust in their results (e.g., be accurate enough to warrant continued use), and scale to the whole enterprise. Additionally, they must be used by management if the culture of the organization is expected to embrace the adoption of such a system (Ehrlich, 2003).

1.5.2   Existing Systems

Systematically identifying experts has been an ongoing research problem for quite some time (Ackerman & Malone, 1990; McDonald & Ackerman, 1998; Lutters, Ackerman, Boster, & McDonald, 2000; McDonald, 2001).

First, the people themselves have been asked to describe their own talents and areas of expertise, but this has demonstrated problems of motivation and incentive, as well as issues involving truthfulness and bias (Fitzpatrick, 1999; Yamim, 1996). Additionally, self evaluation leads to blind spots and the tricky

pre-coordination problem of not knowing who the audience will be. We explain what we do and what we know differently to a colleague in the same field than to someone who does not already have a working knowledge of our own area. We contextualize when describing our skills to others face-to-face, because we can, because we know the audience. When asked to do this for all possible audiences, we stumble.

Alternatively to self-report, the knowledge artifacts that have been previously produced have been investigated and analyzed (Balog et al., 2009). Trying to identify the latent expertise from the documents that are produced and the transactions that have been recorded has been well studied, e.g., reports and meeting minutes (Craswell, Hawking, Vercoustre, & Wilkins, 2001; Balog, Azzopardi, & Rijke, 2006; Balog & Rijke, 2008), email (Campbell, Maglio, Cozzi, & Dom, 2003), and social network analysis (Zhang, Tang, & Li, 2007; Balog & Rijke, 2007). This area is also changing rapidly as we move into technology-mediated social spaces at increasing rates (e.g. corporate installs of social sites and tools like Facebook, delicious, LinkedIn, Twitter). We are producing more artifacts than ever before, which is actually creating a different problem – there is too much. Finding the wheat is proving increasingly difficult and expensive.

Some existing systems include technology that allowed for both self documentation as well as automatic extraction and creation of profiles. The Community of Science's Expertise product allows for scientists in all fields to maintain an expertise profile that can follow them throughout their career, but the fields are self-updated and badly out of date or sparsely populated for

many who have profiles in the system (Fitzpatrick, 1999, 2001). HP's internal Connex directory of experts also allowed for self-description and self-updating (Davenport, 1997; Becerra-Fernandez, 2000). The National Security Agency has an internal staffing and project matching system named the Knowledge and Skills Management System (KSMS) but is based on a custom knowledge taxonomy. Booz Allen Hamilton runs an internal expert skills directory that helps consultants match their expertise with clients' needs (Becerra-Fernandez, 2000). In 2008, Tacit.com sold their expertise location technology, based on automatic profiling from corporate email, and rolled their solution, illumio.com, into Oracle's Beehive collaboration platform. Cameron Marlow's Tagsona is Yahoo's unofficial internal directory that implemented tags and allowed employees to label each other. IBM built Fringe Contacts around the idea that people-tagging is a viable way to categorize "people's skills, roles, and projects in the form of a 'tag cloud'" and was modeled off the earlier IBM work on Dogear, a document tagging system (Farrell & Lau, 2006). Perhaps the most famous of corporate directories, IBM's BluePages house both company controlled information (lines of direct report, past and current projects, contact information) and persona information (controlled/populated by the employee him/herself) (Callahan, 2006a). Most recently Google acquired Aardvark (vark.com) and its question and answer routing technology that is based on semi-automatic expertise profile creation. None of these systems, with the exception of IBM's Fringe Contacts, allows social labeling. They include only self-reported metadata or automatically generated metadata.

I propose another method. I ask, can we not have people talk about what

each other know, and create a new, social, shared knowledge artifact? It should neither be directly derived from the documents produced or by the person being evaluated; it should come from the people around the person of interest. It should come from tacit, social knowledge.

Can we create a knowledge artifact similar to the existing knowledge artifacts, but with a greater ability to encapsulate the here and now and to bend with time? Humans can synthesize a vast amount of context and provide better descriptors and categorize each other in more nuanced ways than perhaps any text mining or latent semantic indexing algorithm can. Even if it is not better, it may provide a different, important perspective not currently harvestable through automated means.

A socially created, shared artifact might quickly adapt to new terminology, new clusters, and see patterns that other systems might take longer to "see." It could be a *new* artifact, one that portends to be the current culmination of knowledge and synthesis. It could be a cutting edge reflection on the knowledge and expertise of a group in the moment.

I want to ask, and then enable, people to help create this new artifact.

1.6   Contextual Authority Tagging

Once we have a social reflecting lens to help us see what a person knows about, it serves as a jumping off point for powerful assessments and assertions. A validated socially robust system of categorized areas of expertise could be the foundation on which to build business and social services.

Can we imagine an ever-available data overlay of expertise? It could be the collective back wall that all ideas get bounced off of before further discussion – a back-chatter that has the opinions you value and need at any time. If it is ever-present and ever-evolving, it could influence nearly every decision we make when we interact with others. It could become the input we need to feel confident. We could eventually feel exposed and vulnerable without it.

It is important to remember that, as we move forward, we do not lose the ability to continue mining all our existing artifacts, documents, and logfiles. These are the raw materials that we use when we generate and manufacture our opinions. The socially constructed representation of one's areas of expertise, the visible version of Wegner's Transactive Memory (Wegner, 1986), would be a new source of information and would only serve to complement what we have already been able to do within the realm of document management (Choo, 1996). Keeping the focus on the people instead of the artifacts they create may better reflect the organizational knowledge inside a group and could greatly reduce the periods of time when new entrants are trying to get their bearings in a new office or managers are trying to assign relevant people to the task at hand.

Contextual Authority Tagging is a proposed technique for expertise location within a group by creating explicit knowledge from the group's individual tacit knowledge about each members' areas of expertise (Nonaka, 1991). This group can be an organization of any size, a loose affiliation of acquaintances or colleagues, or potentially everyone on Earth. For the purposes of this research, the scope of Contextual Authority Tagging will be directed towards the small

and medium-sized working organization and membership. If and when this technique is shown as viable, then a greater scope could be approached, but at this time, some basic assumptions need to be questioned and verified.

Individuals have diverse interests, experiences, and connections with others. Some individuals have a wide variety of areas of expertise with working knowledge across many domains. Other individuals may live a very focused life and have extensive depth of knowledge in one area or two. As sources of information, members of each of these categories of individual are valuable, but in different ways. The Jack-of-all-trades may have insight into how techniques or methods fit together across traditional domain boundaries whereas the deep expert may have encountered a specific subtlety of something that one is beginning to work on and consulting with that person could save one lots of time and money that might have otherwise been wasted.

Knowing which people know which things is key to efficiently leveraging a network of contacts. Routing one's questions, seeking inspiration, and the building of teams each benefit from efficient use of existing mappings of knowledge and areas of expertise. Historically, these types of activities have been hard to commodify or automate. Humans are very good at applying a heuristic for knowing what others know and this research aims to tap into that talent.

Contextual Authority Tagging seeks to create and maintain a mapping of the areas of expertise of a network of individuals. It will do this by having the individuals involved use free text keywords or tags to label each others' areas of expertise. It is explicit and transparent and designed to uncover "reader-generated metadata" rather than "author-generated metadata." Results are

shared back into the group and made visible, and the process is repeated. The resulting product is a weighted list of words associated with each person's areas of expertise. Words are weighted more heavily when more people used those words to tag an individual. Over time, the list, or some subset of the list (e.g., only tags from the most recent 12-month period), would presumably bend and follow the shape of the individual's current interests and knowledge as perceived by the group. Each individual's weighted list would be a specific fingerprint in the multidimensional space created by all possible keywords and could potentially serve as inputs and be used by a multitude of other tools to aid in further decision-making tasks.

CAT is contextual in that each person's fingerprint is unique and relative both to the querier's network and to the queried's network. Limiting whose "votes" count could preempt noisy or "spammy" results. Limiting "votes" with respect to the time they were recorded could prevent "old" or outdated results.

One could imagine future algorithms working in the background, being recursive in nature (similar to Google's PageRank (Brin & Page, 1998) or Kleinberg's HITS (Kleinberg, 1999a)), returning a ranked list of people as weighted by how many other people, who have weight in that domain, "voted" for those listed.

Authority refers to the cognitive authority being granted by the network to each group member (Wilson, 1983). Wilson differentiated between administrative authority (which is obtained by virtue of position or rank) and cognitive authority (which is granted by others based on experience and demonstrated knowledge). The fact that this authority is granted, rather than held "ex

officio," is what makes CAT interesting.

The opinions of one's peers hold interesting collective insights and this technique hopes to tap into this insight and bring it out where both the individual can benefit from her own hard work and expertise and others can more efficiently locate that expertise.

1.7   Research Questions

Contextual Authority Tagging has been conceived and designed to get at two major questions regarding how a group comes to know about its own areas of expertise. The following questions are raised and will be addressed by the following research methodology.

R1. Does CAT work?

    (a) **Similarity** - How similar are a group member's opinion of his/her own areas of expertise and the group's opinion of his/her areas of expertise?

    (b) **Convergence** - How does the similarity behave over time? Do the two opinions converge? If so, how long does it take? If not, is there a persistent gap?

R2. How acceptable is CAT?

    (a) **Comfort** - How comfortable are group members in participating? What are the main factors influencing their comfort level?

(b) **Confidence** - How confident are group members in a system like this? What is the quality of the output of this system? Does this system provide a valid credential? Does this system increase users' trust in one another?

(c) **Usefulness** - What is useful about a system like this? What did participants learn? How would using this system affect participants' decision making?

Latour and Nelson suggest to us that where there is a lack of contention, a social fact will be defined (Latour & Woolgar, 1986; Nelson, 1993). Social tagging phenomena have demonstrated a stabilization of tagging behavior (Russell, 2006; Golder & Huberman, 2005). Together, these suggest the first hypothesis:

H1. As the social fact of what a person knows is molded by the group, a consensus will appear and converge.

The comfort levels of the participants will depend on their surroundings, the familiarity of the task, and their feelings of control:

H2. Comfort levels will increase as the system becomes known and understood. Initial trepidation will be assuaged as the system allows participants to see more of how they are perceived by others.

The warranting principle suggests that we give more credence to information provided by others, rather than information within the control of a particular other (Walther & Parks, 2002; Walther, Heide, Hamel, & Shulman,

2009). Online or offline, information that is known to be easily manipulated is less trusted. Additionally, Delphi-style studies increase the confidence levels of the participants (Rowe, Wright, & McColl, 2005). This leads to the third hypothesis:

H3. Group members will have confidence in this system and exhibit increased trust in one another.

Chapter 2

Literature Review

The following four sections represent core philosophies and research by others that are both interesting and important and are relevant to the work done in this dissertation. In addition to the work mentioned earlier with specific reference to expertise location (Section 1.5.1), the following situate Contextual Authority Tagging within the existing academic literature. Section 2.1 discusses group dynamics and how information flows into, within, and from groups and organizations. Section 2.2 covers identity, reputation, and trust from the perspective of an individual in our newly always-connected, always-on reality, how we understand each other through our past actions and credentials, and how we plan for the future based on that understanding. Section 2.3 is about expertise, what it is, how we think of it, and the artifacts we use to measure it. Section 2.4 looks at the state of the art with regards to tagging, or social labeling, and how it has disrupted the largely top-down hierarchies through which the world has long been described.

2.1 Groups

2.1.1 Knowledge Management

> "If HP knew what HP knows, we would be three times as profitable"
> – Lew Platt, Hewlett-Packard CEO, echoing a former head of HP
> Labs

On the Data-Information-Knowledge-Wisdom (DIKW) hierarchy (Sharma, 2004), knowledge is only one of four things that are hard to define. Culturally defined as *what we know*, knowledge has long been considered to be captured within documents or other containers. A more contemporary understanding would be that "knowledge can only be created dynamically in time" (Newell, 1981, 11). Newell's work goes on to suggest that "knowledge is best conceptualized as an observer-relative attribution: an agent attributes knowledge to an agent observed in order to explain the observed agent's behaviour. It is hardly possible to find out whether the observed agent actually has knowledge as knowledge is dynamically created" (Lueg, 2002, 4).

But knowing that someone knows something is only useful if we can share that information and then act on it. This is most pressing in the organization, where the obligation of the company is to produce a product or service and generate revenue. As such, organizations are constantly struggling with the efficient allocation of scarce resources. They struggle to optimize labor, capital, expertise, knowledge, energy, time, and reporting. Management of an organization, or any part thereof, is tasked with this constant struggle for optimal allocation. Doing it well increases the likelihood of profitability and

customer satisfaction, as well as a better sense of organizational well being and confidence for the next task.

The orchestration of knowing where everything is and how it operates is complex and cannot realistically be handled by a single individual. We have come to depend on each other for knowing what is going on (Johnson, Lorenz, & Lundvall, 2002). Hierarchies develop in organizations; division of labor, workgroups, they all develop because each person can only do so much and it is more efficient if the tasks at hand are separated and conquered individually. This creates another level of management to keep track of the different people doing different tasks.

All that said, organizations are pretty good at this. The field of knowledge management has developed over the course of 20-30 years and seen the rise and fall of management styles and trends. When labor was the most important asset to be managed in the late 19th century, organizational best practices were born out of Scientific Management, or Taylorism (Taylor, 1911). Taylor advocated measurement and optimization on the factory floor and on the assembly line. Later in the early 20th century, statistical methods were applied to the scientific management movement and eventually led to a part of what is now known as Operations Research. Taylor's influence remains part of modern organizational theory and practice, in that we now have departments of work study, personnel, and quality assessment and control in organizations large enough to demand them.

Information machines were introduced into the modern organization in the middle of the 20th century and reformulated the way organizations reported on

their activities. Project status and projections became more specific and finite. We could measure things we could not measure before and with measurement comes an opportunity for further optimization and testing. Fortunes were made at the systems level by shaving a percent here or a percent there and streamlining existing production channels.

When the computers began to generate the bulk of new documents themselves, instead of simply counting what the humans were doing, we entered a new age of storage/retrieval and document management that lasted through the end of the 20th century. We struggled with machine learning and data mining to help us understand and see the patterns in all the documents, logfiles, and artifacts we were producing.

Eventually, Nonaka published his works on the SECI model (Socialization, Externalization, Combination, Internalization) (Figure 2.1) whereby the production of knowledge and value in an organization was actually made up of people working together, learning on the move, and synergizing to produce new knowledge. Nonaka proposed a continuous cycle of knowledge creation between Tacit and Explicit knowledge (Nonaka, 1991, 1994).

Widely cited and commonly accepted today, this cycle is where I plan to couch the approach of allowing individuals to talk about one another's areas of expertise, to bring the tacit knowledge of an organization or group to the surface. Contextual Authority Tagging is a tool that sits along the upper two quadrants of the Nonaka model and helps to externalize a group's opinions on its own expertise.
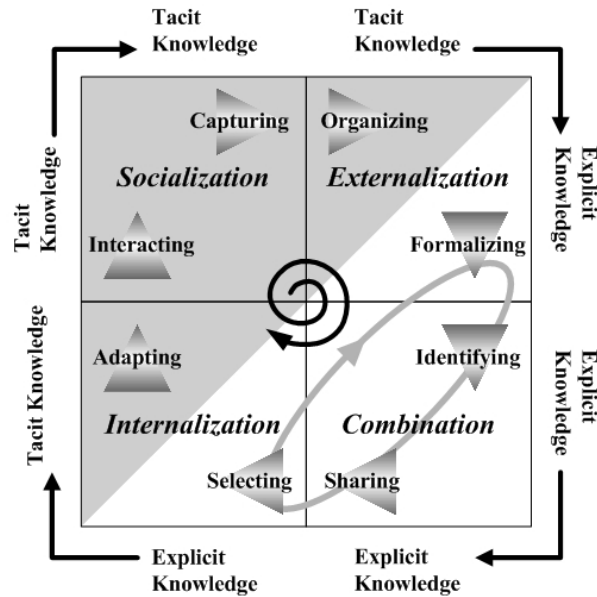
Figure 2.1: Nonaka's SECI Model

Polanyi wrote that "we know more than we can tell" (Polanyi, 1966, 136).
His prime reference is to the ability we have to distinguish faces of others we
know without being able to really describe these faces to others. There is
information in our heads that we cannot communicate with only our sense of
language. His point is that there is information below the surface that we may
not readily be aware of. He calls this information "tacit." Choo says that tacit
knowledge is "personal knowledge that is hard to formalize or communicate to
others. . . . [and] consists of subjective know-how, insights, and intuitions
that comes to a person from having been immersed in an activity for an ex-
tended period of time" (Choo, 1996, 334). By reflecting and becoming aware of
this tacit information, we can begin to describe it and pull it into the realm of
the explicit – that which can be easily transmitted between people and groups.

When suspects' faces are encouraged to be recreated by sketch artists or computer composites, crime victims and study participants both do a much better job of recreating faces they know than without these external aids. Working with the sketch artist, by iterating between description and feedback to what has been drawn, creates better results. The information is there, it just needed better tools to be externalized before it could be effectively shared with others.

Beyond the realm of managing and handling of documents created by humans, knowledge management involves "any process or practice of creating, acquiring, capturing, sharing, and using knowledge, wherever it resides, to enhance learning and performance in organizations" (Swan, Scarbrough, & Preston, 1999). This means that it encompasses other areas such as workspace design, communities of practice, and an understanding of incentives to get people to share their expertise and knowledge.

2.1.2 Community

Communication theory includes the concept of diffusion of innovation. Rogers defined diffusion as "the process by which an innovation is communicated through certain channels over time among the members of a social system" (Rogers, 1962, 5). He defined an innovation as "an idea, practice, or object that is perceived as new by an individual or other unit of adoption" (Rogers, 1962, 12). This movement through organizations or memberships has largely been affiliated with the concept of homophily (McPherson, Smith-Lovin, & Cook, 2001), those people who are similar to one another take on similar behaviors faster and in greater numbers. This leads to a natural imbalance among

members of a group with regards to their experiences and therefore their access to information. This unequal distribution of information leads to a state of having both information-rich members and information-poor members, but not equally across all information domains. The separation of information-rich and information-poor happens roughly independently for each area of knowledge. Every topic will have people who know more about a topic even if they have each had the same training and exposure, since they will group and clump with one another, socially and therefore unevenly. To flatten this distribution, we must be aware of and take into consideration the social aspects of group dynamics. If we can encourage discussion across these normally disparate groups, we may increase familiarity and understanding as well as a better, more even distribution of knowledge.

Groups also self organize within organizations – they create what are known as ingroups and outgroups (Gerard & Hoyt, 1974). The social need for defining oneself via others drives us to seek those who are similar to help define who we are while also defining who we are not. This leads to differentiation among groups. Another effect of this self organizing is the development of a common language within a group (Abrams, O'Connor, & Giles, 2002). Members will define new language for a variety of reasons - these include specialization and opaqueness. An example of specialization is when members within a trade group slip into trade speak fairly quickly as it is a more efficient means to talk about the things they find interesting. Police officers do this while on the job, and may continue the practice when off-duty. An example of opaqueness is when teenagers continually create new slang so that their parents and other

adults seem to realize they are as out of touch with the teenagers as their teenagers want them to feel.

Also among groups, a sense of community forms as people realize they have a shared experience and shared understanding of the world (McMillan & Chavis, 1986). This can be seen in diverse contexts such as the military (boot camp), at corporate retreats (trust fall), and summer camp (pitching a tent). To know that someone else has been through the same experience means that you know a little bit about what they are and what they have been through, things that make them who they are. This shared understanding makes the members of a group more intimately aware of each other's perspective.

When the members of a group are interacting mostly online, this sense of community and shared understanding has been called "ambient intimacy". People who are not sharing physical time together have reported a sense of closeness to those they are connected to, at a distance, through the mediating technologies of the Internet (Reichelt, 2007). This matches earlier group work on network proximity that shows members who are more close are more exposed to social information and more likely to be influenced by that information (R. E. Rice & Aydin, 1991). The information flows are directly related to the density of an actor's network such that members who are tightly connected to many others are more heavily influenced by their peers.

When a community exists in an online space, certain organizational dynamics are made available. One of these is the freedom from "specializing roles by geographic location" (Rosedale, 2009). When a team's communication channels are composed of data moving across a network rather than through the

air in a physical space, the team can be widely distributed. When the communication channels are sufficiently robust and fast, and "as communication technology makes transparency cheaper, the need for central control drops" (Rosedale, 2009). Teams can be composed from the most talented, most compatible people from anywhere in the world, not just from the talent nearby or on hand. Managing those teams can prove more complicated, but the talent and demeanor of the team can overcome the added complexity.

Community of Practice theory holds that there is value beyond the tacit – the very existence of a social order suggests there is more to knowledge than the codified explicit and uncodified tacit (Duguid, 2005). Duguid says that because we are social creatures, we create information of a social nature. It is not explicit or tacit (as the SECI model suggests), but rather, social. I would suggest that the sociality of information is a separate facet or spectrum of the information, and not a separate type altogether. Communities that share information due to the fact that they are a community have identified a useful outlet for some of their collected tacit information.

2.1.3 Incentivization

The incentives involved in encouraging community members to contribute to the common goal are many. In large part, membership in a community or organization is driven by self-serving motivating forces. Barnard summarized this position in 1968 as a basis for his seminal work, *The Functions of the Executive*:

> The contributions of personal efforts which constitute the energies

of organizations are yielded by individuals because of incentives. The egotistical motives of self-preservation and self-gratification are dominating forces; on the whole, organizations can exist only when consistent with the satisfaction of the motives, unless, alternatively, they can change these motives. The individual is always the basic strategic factor in organizations. Regardless of his history or his obligations he must be induced to cooperate, or there can be no cooperation. (Barnard, 1968, 139)

Barnard goes on to explain that organizations can induce participation (or membership) either through objective incentives or persuasive methods. What he calls objective incentives include both specific objective incentives (material items, physical conditions) and general incentives (communion, associational attractiveness, participation) while his persuasive methods are called such because they affect the subjective state of mind of the member in question (creation of coercive conditions, rationalization of opportunity, and inculcation of motives). Barnard says that the objective incentives are used mostly by industry and industrialized organizations where monetary consideration is stable and normative and persuasive methods are used predominantly in religious and political organizations. He is also careful to point out that both types are used in all organizations but that he had observed the distributions above across many organizations (Barnard, 1968, 141).

When a new group forms, participation may begin with enthusiasm, but as roles and norms settle out, keeping the energy and contributions at a high level becomes harder. Keeping membership motivated and interested requires keeping them incentivized. Clark and Wilson (1961, 134) suggest that incentives can be categorized as either 1) material (tangible and/or economic and valuable to the membership), 2) solidary (intangible, social, involving status, and

unrelated to the goals of the organization), or 3) purposive (also intangible, but related to the goals of the organization). Knoke finds that individual members of associations are motivated by three types of influence: 1) rational choice (cost/benefit analysis of expected utility), 2) affective bonding (emotional attachment to other members of the group), and 3) normative conformity (doing what others like themselves are doing) (Knoke, 1988). The determinants for participation in volunteer organizations and activities have been found to be "larger context (territory and organization), social background and role variables, personality traits, attitudes, and situational variables" (Smith, 1994, 256).

Overall, motivation to belong and participate seem to revolve around economic incentives, social incentives, and political incentives – a very familiar triple of considerations. Related to Clark and Wilson's third category of purposive incentives, Elinor Ostrom recently won the 2009 Nobel Prize in Economics for her work (Ostrom, 1990) suggesting that common goods *can* be managed with purposive collective oversight, and under certain circumstances avoid the classic *Tragedy of the Commons* that demands government oversight or private ownership to manage resources (Hardin, 1968). Ostrom writes about global trust and global payback in the sense that members give back to the organization because they felt obligated through what they had received, and this payback could be at a value much greater than the value the member originally received.

2.1.4   Collective Intelligence

A definition of collective intelligence comes from Lévy, just as the advent of the web took place (originally published in French in 1994), "It is a form of universally distributed intelligence, constantly enhanced, coordinated in real time, and resulting in the effective mobilization of skills . . . No one knows everything, everyone knows something . . . " (Lévy, 1997, 13). Shortly thereafter, Heylighen stated that "collective intelligence is defined as the ability of a group to solve more problems than its individual members" (Heylighen, 1999, 253). More recently, collective intelligence has been described as a *Wisdom of Crowds* (Surowiecki, 2004), *Smart Mobs* (Rheingold, 2002), or more entertainingly, *Here Comes Everybody* (Shirky, 2008a). These different ideas sound very similar, but have a few distinct differences. Surowiecki thinks that the participants need to be diverse, act independently and without central control, and their wisdom will present itself in aggregation. Rheingold suggests that the intelligence of a group comes from the vast array of within-network linkages given our newfound digital connectedness. Shirky says that we do not need formal organizations to help us figure out how to act as groups – that we can do that on our own now. Lévy saw us moving into the fourth of our social spaces – from Nomads, to the agrarian Territory, to the commoditized Property, and now into a shared space of Knowledge.

Another way of thinking about collective intelligence may be as distributed cognition. Distributed cognition suggests that the groups of people working together towards a coordinated end are communicating and sharing information in a socio-technical system (Hollan, Hutchins, & Kirsh, 2000). The creation

and transfer of artifacts and signaling within these systems allow the group to succeed at their task. When knobs are turned and notes taken and commands relayed, the members of the group are part of a collective process, a collective cognition, that helps the group learn and understand. Distributed cognition also plays out over time, as the culture and environment of the group share information with group members that come later.

Related to the environment, Activity Theory suggests that the tools we create come directly from and map to our mental processes. When tools are created within a group, to manage the repeat work and save human effort, they are manifestations of the cognitive effort and the shared culture of the group (Vygotsky, 1978). A direct result is that the shared artifact of the tool and the work it represents become encapsulated. What had been in the minds of the group can be communicated with a shared understanding of what the tool does. This allows the human mind to work on new problems and move on to encapsulate new knowledge. It also allows complex procedures and activities to be broken down into their requisite steps and understood in pieces (Nardi, 1995).

Some work being done at MIT has led to a mapping of the "genes" that are part of collective intelligence systems. These genes include the answers to four main questions (Who, Why, What, How) and so therefore consist of issues around Staffing, Incentives, Goals, and Structure/Processes (Malone, Laubacher, & Dellarocas, 2009). By breaking down the parts of what makes collective intelligence systems tick, they hope to be able to then understand and manipulate how these genes fit together within networks of people.

The underlying theme throughout this diversity of opinion and work is the idea that the communication technologies that we have built up and continue to integrate into our everyday lives are essential to a burgeoning collective intelligence. It is through this "synergy between human and machines" where "machines are the enablers: they store and remember data, search and combine data," and where "people learn by communicating with each other" more efficiently via those machines that we are capable of communicating across time and space with such increasing power. That there are "different roles for people and machines" and that "people learn by communicating with each other" are both central themes in my future work (Gruber, 2008). Computers are good at connecting, storing, and counting. Humans are good at socializing and finding meaning – they are the "producers and the customers" in any system that resembles having some collective intelligence.

Additionally, I believe that to understand collective intelligence, we must realize that the frameworks and the research named above are predicated on three key factors that must be considered and understood: that any intelligence comes about because of an underlying network and that network's own network dynamics, that collaboration and awareness must exist between the participants of the network, and that software is the enabling tool to help us quantify and then analyze what is happening.

2.1.5   Networks and Network Science

Network science has recently come into its own as a discipline. With the publishing of books by Barabási in 2002 and Easley and Kleinberg in 2010,

the study of networks and the interconnectedness of the systems in which we live have found themselves front and center of the discussions of how Web 2.0, among other things, touches most of our lives in profound ways (Barabási, 2002; Easley & Kleinberg, 2010).

Networks comprised by social systems largely contain Pareto distributions of connections between the items being linked. This means that roughly 80 percent of the links point to 20 percent of the nodes. More generally, this has been found to be the distribution of time spent on decision making in meetings, paper citations among academics, and cities served by the airline industry. Networks where human activity is involved usually exhibit these power law, or Pareto, tendencies and can be modeled with logarithmic techniques that are strikingly predictable. Plotted on a logarithm-logarithm chart, power laws appear as a straight line. Other related phenomena include Zipf's Law (the distribution of words in the English language is linear on a log-log plot) and Moore's Law (the number of transistors on a chip doubles roughly every two years).

Knowing how networks are constructed and how they grow has also allowed researchers to predict the effects of these networks on other things (e.g., creating network effects). In the broadcasting medium of radio and television, Sarnoff's Law suggests that as each new entrant appears on the network (e.g., a consumer purchases a television), the increase in the value of the network is proportional to that single viewer. This means that the network's value is equal to the total number of viewers, or n (Reed, 1999).

Bob Metcalfe observed that in a bidirectional communications network,

where he worked in industry, a new entrant into the network increased the number of connections, and therefore the value of the network, in a quadratic way, as the square of the number of "compatibly communicating" network nodes (Simeonov, 2006). Using fax machines as an example he pointed out that a single fax machine had very little value, as there was no one to communicate with. A second fax machine added a single link. A third fax machine added two links, and a fourth, three more links. This formula of $n(n-1)/2$ is dominated by the n^2 term and can therefore be modeled as simply n^2 and has come to be known as Metcalfe's Law (Gilder, 1993).

More recently, Reed suggested that within a network there are actually many more connections being made than between individuals. In addition to connections being made between individuals, groups themselves are making connections. Group Forming Networks have additional network properties and should be measured differently. Reed's Law says that the value of the network itself can be modeled with an exponential or geometric formula on the order of 2^n which grows much faster as additional nodes are added (Reed, 1999).

There have also been suggestions that Metcalfe's Law is overstated since not every participant in a network can actually communicate with every other network member. Odlyzko and Briscoe have suggested that a more practical measurement of the value of a network is $n * log(n)$, but have not produced any formal proof. These are mostly abstractly useful models that predict an upper bound on "value" (Odlyzko & Tilly, 2005; Briscoe, Odlyzko, & Tilly, 2006).

Additionally, much work has been done with regards to how the networks

themselves are comprised and the interconnectedness that has been measured. Granovetter described the power associated with weak ties in our social networks. He said that most of the value in our networks come from our loose affiliations, or weak ties, and not the ones that we consider strong (Granovetter, 1973).

Kleinberg wrote about hubs and authorities as indicators of influence in the networked world (Kleinberg, 1999b). Hubs are defined as those nodes that point to many other nodes. Authorities are defined as those nodes that have many nodes pointing to them.

Formalizing and generalizing what Milgram called the Small World problem in the 1960s (Milgram, 1967), Watts and Strogatz described with remarkable clarity how deeply interconnected we are and how networks usually have a "giant component" where the majority of group members are well-connected (Watts & Strogatz, 1998). The areas of social capital and reputation research have also benefited from our understanding of network theory. Social capital theory, pulling together both Kleinberg's hubs and authorities model with the strong and weak tie work of Granovetter, shows how bonding and bridging capital allow us to model our relationships in a predictably economic-like way (Lin, 1999).

Understanding the implications and findings of this network science research is critical when planning or building out new network features and capabilities. My hope is that Contextual Authority Tagging can make some qualified, safe assumptions about network topology and connectedness when

evaluating the results of having people tagging other people's areas of expertise. In short order, the hope is that findings within smaller groups can be eventually claimed to be more generally applicable via *Small World* and *Strength of Weak Ties* theory.

2.2 Identity, Reputation, and Trust

"I am a part of all that I have met" – Alfred, Lord Tennyson, *Ulysses*

2.2.1 Identity

2.2.1.1 Offline

Identity is something that has fascinated and puzzled mankind for generations. Descartes wrote "I think, therefore I am." He was working out the existential questions of what it means to be human – of what it means to consider the world and all that is in it. But he was also working through his own relationship with the world. This sense of relationship, of belonging, is something that each and every one of us comes to question as we come into our own and as we change throughout our lives.

Goffman's seminal work, *The Presentation of Self in Everyday Life*, conjures a metaphor that has served most of sociology, psychology, and many other social sciences very well over the years (Goffman, 1959). His dramaturgical metaphor of the theatrical performance where we each wear different masks and present to different audiences on different stages is extremely satisfying. He writes that:

> The stage presents things that are make-believe; presumably life presents things that are real and sometimes not well rehearsed. More important, perhaps, on the stage one player presents himself in the guise of a character to characters projected by other players; the audience constitutes a third party to the interaction - one that is essential and yet, if the stage performance were real, one that

would not be there. In real life, the three parties are compressed into two; the part one individual plays is tailored to the parts played by the others present, and yet these others also constitute the audience. (Goffman, 1959, Preface)

As we work and play with one another, the ability to move and morph between different presentations of self are essential. We cannot be the same person all the time to all people. We dictate our actions on a huge number of variables but the big ones include who the audience is and our desired outcome; the means and the end. Both matter, and neither really justifies the other, but both are essential in determining how we act. Social identity theory (SIT) suggests that our sense of self comes from two places, our social and our personal identities. The personal comes from our unique characteristics, whereas the social comes from our shared cultural and interest groups (Tajfel & Turner, 1986).

With regards to the presentation of self to others, there is a distinction to be made between presenting a persona or version of oneself and presenting an entirely different version of oneself which is outright deception. One is a natural product of us having social circles and the necessary human reaction to interacting with a diversity of social connections in different contexts. The other is a legally abhorrent means of fraud or impersonation.

The traveling medicine man of the 19th century (huckster) was a rare successful embodiment of this latter phenomenon (McNamara, 1971). In today's connected world the ability to defraud people in consecutive towns has largely gone away because it is extremely hard to reinvent oneself without appearing strange to normative observers ("You have no prior work experience?", etc.).

To attempt a reinvention of this type means that the stakes are very high and one does not necessarily care about the ramifications. It is a Hail Mary pass with an unlikely positive outcome for the defrauder.

In the physical world, our laws are set up to allow the social masks we put on for different audiences and to discourage and punish the outright deception and fraud. In the online or mediated space, it is much harder to police this distinction. We do not yet have the mechanics to reliably know the identity of the other party in a transaction. We certainly are not prepared with thousands of years of practice with these tools – as we have been in the offline, face to face world.

2.2.1.2 Online

Our understanding of our sense of self is being challenged in the online space. As we continue to hurtle forward into a fully networked world, where communications are being mediated through electronic means to a greater extent, we have a new, different set of attributes and rules to play by. This new stage, the mediated presence, affords new behavior and norms, but we do not yet have best practices or agreed upon senses of what they mean. We live in a new age of "cheap pseudonyms," where it is easy to start over, where there is no history attached to a new account (E. Friedman & Resnick, 2001).

Across cultures, we struggle in the physical world, but at least we can be embarrassed, confused, or angry and the effect is limited by time and place. With online mediated identity, we are often presenting the same self to many more people. There is a loss of context of time and place. Networked publics,

modeled after Negroponte's descriptions of bits versus atoms (Negroponte, 1995), are creating havoc with our known patterns of presentation and normal socially aware behavior (boyd, 2008). These new networked publics have the following properties:

- *Persistence*: online expressions are automatically recorded and archived.
- *Replicability*: content made out of bits can be duplicated.
- *Scalability*: the potential visibility of content in networked publics is great.
- *Searchability*: content in networked publics can be accessed through search.

In light of these new publics, the future norm-brokers, the youth of today, have already begun to behave accordingly (Stutzman, 2006). The sharing practices of youth are more pronounced than adults – they use more of the privacy tools available. The youth do not understand how all the technology works, but they do realize that if unwanted people can see into these online spaces, then they lose some of their sought-after and ever-elusive autonomy. Parents are concerned about their children's safety, both online and offline. Because of this fear and the (over)pre-cautious behavior of the parents, the ability to play as a child is going away, and with it the ability for youth to experiment with who they are and what they want to project to the world (Skenazy, 2009; boyd, 2008).

Daniel Solove paints the future of privacy as a balance between our published selves and the interests of the organization or firm. He feels we have largely given up our privacy today and the future will only see more of the

same unless we curb, through law, the tide of this loss of control (Solove, 2007). If the law does not step in, Solove feels we will capitulate even more of our power over our privacy and the data we generate to corporate interests, who, by law, have profit as their primary goal.

In the ramp up of user-centered technologies that are being developed to help us navigate the Internet as people, Kim Cameron has come to the fore in helping to define and construct what an Identity Metasystem would look like. In 2005, he published the *7 Laws of Identity* that have to be met by any system claiming to handle digital identity in a user-centric fashion (Cameron, 2005a).

1. *User Control and Consent*: Identity systems must only reveal information identifying a user with the user's consent.

2. *Minimal Disclosure for a Constrained Use*: The identity system must disclose the least identifying information possible, as this is the most stable, long-term solution.

3. *Justifiable Parties*: Identity systems must be designed so the disclosure of identifying information is limited to parties having a necessary and justifiable place in a given identity relationship.

4. *Directed Identity*: A universal identity system must support both "omni-directional" identifiers for use by public entities and "uni-directional" identifiers for use by private entities, thus facilitating discovery while preventing unnecessary release of correlation handles.

5. *Pluralism of Operators and Technologies*: A universal identity solution must utilize and enable the interoperation of multiple identity technologies run by multiple identity providers.

6. *Human Integration*: Identity systems must define the human user to be a component of the distributed system, integrated through unambiguous human-machine communication mechanisms offering protection against identity attacks.

7. *Consistent Experience Across Contexts*: The unifying identity metasystem must guarantee its users a simple, consistent experience while enabling separation of contexts through multiple operators and technologies.

These seven laws will be the bedrock of good identity systems that get built in the next few years. Erring on the side of minimal and justified disclosure based on the user's consent will be a welcome change from the types of systems and practices we see in place today with regards to data handling and management in most places that have to worry about this kind of personal information. Having options for the users, with regards to service providers, portability, and identifiers keeps the person whose data this is, in charge. Requiring humans to be a part of the mix and keeping things consistent are essential as well if anyone other than the engineers who have been thinking about this type of technology for years is expected to use it or like it.

These laws look past the severe hurdles of corporate politics, legal wrangling (domestic and international), and technological feasibility and compatibility. These facts are not faults, though, but rather, features. Having the big picture in sight allows for a common frame of reference as we move forward with a rather large, rather impossible-looking task. But it will happen, because it is important.

2.2.1.3   Infrastructure

Many community-led efforts have been making progress on the Identity Metasystem vision. Before the term had been coined, technologies such as Shibboleth, x509, LDAP, SSL, PGP/GPG, and cryptography more generally were

being used successfully in the marketplace to help identify and credential users and transactions across the internet. Most credentialing had not been done on the World Wide Web, as the interfaces were too raw and bringing the user into the transaction usually only ended in confusion and frustration.

More recently, newer technologies and projects have started to dot the landscape and make real infrastructure and user interface progress; information cards, the Liberty Alliance, OpenID, SAML, Heraldry, Higgens, Pamela Project, OSIS, and Bandit to name a few. An explosion of both consumer awareness and interoperability are leading the efforts towards widespread identity infrastructure. In a few years, we will begin to see robust reputation and trust systems built on top of a reliable and credible identity layer and metasystem. There are too many large players and too much money to be made to not expect a convergence of technology and thinking in this space. Consistent, interoperable standards around identity will allow for new markets to open up and innovation to push the edge of what is possible.

On the individual level, there are a few coping mechanisms in seeing our offline and online personas begin to blur. Michael Wesch has called it "context collapse" (Wesch, 2008a, 2008b), and danah boyd has referenced "context collisions" in her discussions around Facebook Friends or Friendsters, suggesting that these are not your *real* friends (boyd, 2008). Facebook itself promotes the fact that your connections on their site are of real IDs – that there is social value in real existing connections, and not just people we have met online. Much social research suggests that this is true, but others say that with enough time and energy, online relationships are no less important or real than those

we groom offline as well. Additionally, research has shown that those people who share more in public are more willing to let their personas overlap and verify that they are the same person (Russell & Stutzman, 2007). Some people try to keep multiple personas online, and to keep them separate from one another. This is largely futile in the long run (Novak, Raghavan, & Tomkins, 2004; Krishnamurthy & Wills, 2009), but there are reports of people going to remarkable lengths to keep their public and private, or work and family lives separate (i.e., Belle de Jour) (Knight, 2009). Of course, the counter example is that we do not know about the cases we do not know about – the successes.

2.2.2 Reputation

Upon identity, reputation can be constructed. Without identity, reputation and trust are much harder, if not impossible to employ. Reputation and trust are concepts that go hand in hand in the literature. One cannot read very deeply without finding either competing definitions or overlapping patches of claimed groundwork.

Phil Windley and his students have most recently produced a cogent set of the twelve elements, or characteristics, that any system used for reputation should include (Windley, Tew, & Daley, 2007). Each is a distillation of earlier research and it currently stands as the single most concise list of attributes of reputation. Contextual Authority Tagging, I think, successfully includes or employs all twelve.

- Reputation is one of the factors upon which trust is based.
- The expectation of future reciprocity or retaliation creates an incentive

for good behavior in the present.

- Reputation is personal.

- Reputation is a currency.

- Reputation is narrative.

- Reputation is based on identity.

- Reputation is based on verified claims and transactions.

- Reputation is based on opinion (indirect information from other witnesses).

- Reputation exists in the context of community.

- Reputation exists in a particular context.

- There is a natural tradeoff between reputation and privacy.

- The quality of a reputation calculation should be regularly assessed.

Sabater and Sierra compiled a wide list of computational models that attempt to calculate or otherwise measure trust and reputation between actors in a network (Sabater & Sierra, 2005). This computer science work largely made it clear that there are many measuring sticks currently being used to write software and that, in large part, they are not measuring the same things yet – making it hard to compare the algorithms. Sabater and Sierra classified the algorithmic models across seven facets. Algorithms were classified by conceptual model (game theoretical or cognitive), information source (direct, witnessed, sociological information, and prejudice), visibility (local or global calculations), granularity (contextualized or general calculations), behavior assumptions (lying), boolean/continuous measures, and reliability information. Sabater and Sierra notice that the game theoretical paradigm is dominant,

probably due to the backgrounds of those doing the current research, and that there is more need for input from psychology and sociology researchers in this area. As the models become more complex, the current batch of game theoretical models do not continue to perform as well as with earlier, simpler e-commerce models. Additionally, contextualized models have not been used as much recently, as they are more complex to model. Again, these global trust and reputation scores work well with simple models, but begin to show their limitations when presented with more social problems and decision making tasks.

One of the facets that is implied in the earlier discussion is that of collusion. When people or agents work together, under cover of deception, to further a particular goal, they are lying together. Lying, itself, is an important piece for this research, for when many of these identity relationships have historically happened between people in the physical world they have a human sense of whether the person standing in front of them is the same person who was standing before them the day before. Identity, in this sense, is stable and assumed in the physical world in which we have so much collective history (and practice). And with this, the assumption is that the person who will be standing before you tomorrow, claiming to be your friend, is still the same person.

Online, this assumption of stable identity is not as strong. We do not currently have a good, strong means of identifying actors on the global network with any confidence. There is ongoing work in the Identity community (see Section 2.2.1) to help solve some of these issues, so that our online interactions

can begin to function a little more like our offline interactions – with confidence and more stability over time.

At the societal level, we do have means of assessing reputation. Today's banks, credit rating agencies, mortgage companies all work together to share information about their customers so they can make informed opinions about one's future likelihood to be financially solvent and capable of paying back their loans. In large part, one's creditworthiness now sits upon a single computed value, one's credit score. It is worth noting here that while one's credit score is available to you and others (at a cost, of course), the algorithm (or algorithms) used to determine that score is not. There is little transparency into a process that holds a vast amount of control over one's economic reality in modern life.

2.2.2.1 Source Selection

Another body of research is centered around source selection among peers in a work environment. Source selection is defined to be the process or decision-making that one goes through in deciding which sources of information to use for the task at hand. Many studies over the years have crystalized the notion that source selection is based primarily on source accessibility and source quality. The definitions of these two concepts have been up for debate as well, as accessibility has been defined in terms of physical proximity (Pinelli, Bishop, Barclay, & Kennedy, 1993), comfort with a source (Fidel & Green, 2004), effort involved (Marton & Choo, 2002), as well as simple availability (Vancouver & Morrison, 1995). An interesting social effect is found around the possibility of appearing incompetent when seeking information from a trusted, quality

source (Cross & Borgatti, 2004). Through think aloud exercises, recent research suggests that source quality is the most dominant factor when making these source selection decisions (Woudstra & Hooff, 2008).

If source quality is the primary factor determining source selection, then aiding the effective identification of quality sources follows as an important factor. O'Reilly found that quality sources are perceived to be relevant, timely, specific, and accurate (O'Reilly, 1982). Later, authority and expertise and trust were most frequently cited as criteria for acceptance or rejection of an information source (Nilan, Peek, & Snyder, 1988; Halpern & Nilan, 1988). A few years later, work on information quality and cognitive authority was made more explicit and showed that the credibility of an information source or document relied heavily on the belief that the source was coming from a reliable place (based on reputation, prior work, and apparent authenticity) (Rieh, 2002).

As we continue to struggle with source selection, determining the veracity of a claim of authenticity and vetting the credibility of the provenance of a source become arguably more important than the information held within the source. Finding good data starts with finding good sources of data. Of course, this depends on finding good data on good sources of data. As they say, "It's turtles all the way down" (Hawking, 1988).

2.2.2.2   Social Capital

Social capital is a loosely defined, cross-disciplinary idea spanning sociology, economics, medicine, political science, and psychology. It has been studied

in all of these fields, and the definitions used vary widely depending on the context of the question being examined and the backgrounds of the researchers involved. A definition that should be agreeable to most of the researchers in this area is that social capital "is about the value of social networks, bonding similar people and bridging between diverse people, with norms of reciprocity" (Claridge, 2004; Dekker & Uslaner, 2001). Claridge compiled a table of over twenty definitions of social capital and found that they could be grouped by whether the definitions focused on internal and external relationships, bonding and linking relationships, or both of these types (Claridge, 2004).

Seeing that social capital theory involves a human element, and not a necessarily economic element, researchers have a hard time quantifying and therefore measuring social capital. Others have shown that social capital is impossible to measure directly and that proxies must be used in any attempts at empirical analysis (Collier, 2002). The models that have been developed over the years each have a specificity to them belying the underlying research area and for the most part cannot be said to be comprehensive (Claridge, 2004). That said, social capital theory is widely held as having value and is being incorporated into more and more of the models being proposed that govern understanding of human behavior and decision-making (economic and otherwise). Social capital theory brings together important sociological areas of research including social support, social cohesion, and integration theory (Requena, 2003).

At some point during a group's bonding, group participants have been shown to have a greater connectivity and cohesion with one another (Baker &

Dutton, 2007). Even if the group was part of the control in an experiment, benefits afforded to them by fellow group members begin to be paid back at a greater rate (Yamagishi & Kiyonari, 2000).

The organization, or firm, has also presented itself as a rich environment in which to study social capital, as it is "conducive to the development of high levels of social capital" (Nahapiet & Ghoshal, 1998). Contextual Authority Tagging leverages the environment of the workplace or organization "as a social community specializing in the speed and efficiency in the creation and transfer of knowledge" (Kogut & Zander, 1996).

Lin was the first to take social capital theory and formally apply social network analysis. His work placed social capital into the realm of the measurable and potentially granted the power of causality to some of the social linkages within a network (Lin, 1999). He also spoke about how reputation is an indication of social gain and that this reputation is an aggregation of good will, or, social capital (Lin, 2001).

Moving the discussion of social capital to the web, Uslaner concluded that the Internet is "neither a dark and threatening place nor a grand intellectual and social commune" (Uslaner, 2000). People will remain people and the network is a tool through which they will continue to work, play, trust, and distrust each other. The network does not remove the social from social capital, but because it does hamper the non-verbals, the highs are higher and the lows are lower. We are still on our own to determine the intentions of the person on the other computer.

2.2.2.3 Shared Understanding

Humans have a unique ability to put themselves in the minds of others. Dunbar's Theory of Mind suggests that most humans can place themselves into recursion, or model, three or four levels deep before getting confused (Dunbar, 2004b). This means that we can think about another person's thinking about a third's thinking about a fourth. Dunbar pointed out that Shakespeare regularly worked on the sixth level of recursion: Shakespeare as the writer must intend that the audience believes that Iago intends that Othello supposes that Desdemona loves Cassio, who in fact loves Bianca.

This ability to imagine what others imagine also means that we are capable of manipulating those perceptions in others. We act in calculated ways and we use gossip as a social cue to control the behaviors of others. Dunbar writes that we can control the number of free riders in our midst via social grooming and gossip (Dunbar, 2004a). This assumes stable identities and, with them, we are very effective at limiting those who cross norms and break social standards.

Organizational psychology has developed a tool called "360-degree feedback" or "multisource feedback" or "multisource assessment." First used by the Germans prior to World War II, multisource feedback has steadily gained in popularity over the last 60 years. As an evaluation method, it is now in very wide use in organizations to evaluate worker performance from all perspectives – peers, bosses, subordinates, and others (Fleener & Prince, 1997). Rather than only receiving appraisals from direct reports or from those above, multisource evaluations provide multiple perspectives on a worker's output or

performance. Multisource feedback may also include self evaluation. Most believe that this style of feedback allows for a more full, more nuanced view of an employee to be obtained during review, but consensus has not been reached on whether this knowledge about the worker effects real change (Seifert, Yukl, & McDonald, 2003). Being evaluated by others also brings a cognitive load that forces the participant to see others' perspectives – an invocation of Theory of Mind is part of what makes going through the process hard for the employee.

The Johari Window (See Figure 2.2), named after its creators, is another device that has come into its own as a useful representation of the level of awareness of interpersonal relationships between people (Luft & Ingham, 1955; Luft, 1961). When evaluated, a group member or researcher categorizes the information being revealed about the group member. The four panes of the window are as follows and represent the different types of information about a person.

- *Quadrant I*, the area of free and *open* activity, refers to behavior and motivation known to self and known to others.

- *Quadrant II*, the *blind* area, where others can see things in ourselves of which we are unaware.

- *Quadrant III*, the avoided or *hidden* area, represents things we know but do not reveal to others (e.g., a hidden agenda or matters about which we have sensitive feelings).

- *Quadrant IV*, area of *unknown* activity, where neither the individual nor others are aware of certain behaviors or motives. Yet we can assume their existence because eventually some of these things become known, and it is then realized that these unknown behaviors and motives were influencing relationships all along.

The Johari Window may be such a useful tool and graphical model for
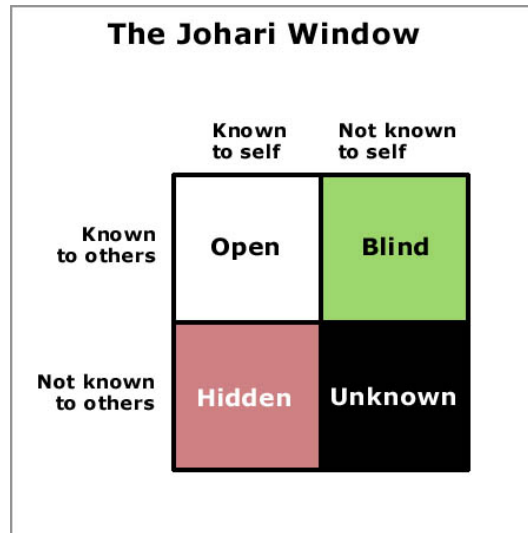
Figure 2.2: Johari Window of Interpersonal Relations

an organization since we know about social comparison theory. Humans are constantly sizing each other up and trying to make sure they fit in with their social surroundings (Festinger, 1954). However, our ability for "social projection" (Gerard & Orive, 1987, 171), that of estimating others' attitudes about oneself, has often been shown to be "dichotomous" with ones own attitudes about oneself. If not kept in check, this dichotomy can lead to a false sense of consensus and agreement between coworkers and group members (L. E. Rice & Mitchell, 1973). We need to be reminded of the shortcomings of our own ability to evaluate social situations so that we can continue to re-evaluate and incorporate all perspectives.

One method to detect a dichotomy among a group, or a lack of "Shared Understanding" is that of *The Squirm Test* (E. E. Kim, 2009):

> The Squirm Test is a simple tool for measuring Shared Understanding. Take a team of people working on a project together. Have

them sit in a circle and on their hands.

Ask someone to stand up and briefly explain what the team is working on, what are the challenges, and what's [*sic*] the plan moving forward. No one is allowed to say anything unless they are standing. Once that person is finished talking, have the next person stand and go through the same exercise. Repeat until everyone has had a chance to speak.

The more people squirmed while others were talking, the less Shared Understanding you have.

Applied to people, instead of projects or teamwork, *The Squirm Test*, or a test like it in form and function, could be a good indicator of shared understanding about a person's areas of expertise or knowledge.

Recent studies by Vazire and Mehl go further. While there may be blind spots and missing information among groups in a social setting, evaluations of each other cannot be comprehensive without the input from all parties. "There is no single perspective from which a person is known best and that both the self and others possess unique insight into how a person typically behaves" (Vazire & Mehl, 2008, 1202). We need everyone to participate in thinking about one another if we hope to fully capture the social knowledge within a group or organization.

2.2.3 Trust

The new, digitized transparency is one major means of facilitating deals between people who do not know each other. (Etzioni & Bhat, 2009)

Like a few of the topics in this literature review, there are both philosophical definitions as well as modern, computer-related definitions and research

regarding trust. Baier wrote about *Trust and Antitrust* in 1986 and is commonly referred to in matters of defining trust as we understand it in a modern society (Baier, 1986). Citing works from Aristotle, Hume, Plato, Locke, and Hobbes, she points out that largely, philosophical debates regarding trust have been rather sparse. In addition, Baier argues that most discourse is on trust between rational peers not of unequal power – objective and dispassionate discourse among those equal in a social hierarchy. She finds this to be a shortcoming of the literature and the discussion, but at the same time, while pointing out that the majority of relationships in the world are not of this type, rational peers serve as the best place to get at the moral questions regarding free will and trust.

She says that trust could be evaluated in terms of betrayal rather than just reliance. "One leaves others an opportunity to harm one when one trusts, and also shows one's confidence that they will not take it" (Baier, 1986, 235). The more information one passes to someone else, the more power one grants to them – power to do both good and bad. When someone is granted the power to betray but with the confidence that they will not betray, trust has been conferred. To this end, trust can be measured as a function of confidence and power.

As well, "trust is much easier to maintain than it is to get started and is never hard to destroy" (Baier, 1986, 242). Along with reputation, trust can be quickly reduced to doubt and questioning by a breach of confidentiality, morality, or follow-through. Axelrod said in 1984 that trust allows us to give value to the "shadow of the future" today (Axelrod, 1984). All our information

flows have "historical residues" moving forward and should be considered with every decision we make (Fisman & Khanna, 1999).

Research on trust within information science has tended to focus on the relationship between the user and the document or the information that is being interacted with (Kelton, Fleishmann, & Wallace, 2008). Other researchers complain that trust can only be between two people and not computers or information sources (Solomon, 2000; B. Friedman, Peter H. Kahn, & Howe, 2000). Kelton defined four levels of trust related to the scope of the players involved (Kelton et al., 2008, 364):

- *Individual*: Personality characteristic
- *Interpersonal*: Social tie directed from one actor to another
- *Relational*: Emergent property of a mutual relationship
- *Societal*: Feature of a community as a whole

This scoping of trust creates a clean perspective into existing research and keeps separate the vastly different working definitions of trust. Kelton explains that the Individual level is largely defined by the Psychology literature where trust is defined more as a predisposition *to trust*, a personality characteristic. Interpersonal contains the most work and characterizes trust largely as a measure of expectation or confidence in another actor's future behavior. This encompasses the computer science and modeling work mentioned below. The Relational level is smaller since it assumes a mutual relationship, and Uslaner and Lin's work on social capital fall into the Societal category.

While Baier speaks mostly about trust between consenting, rational adults

with the capacity to continuously make decisions about past, current, and future transactions, the literature within computer science is largely constructed around the most primitive model of a relationship, that of one software agent to another. Trust in this regard is interesting as the exhibited behavior can be deterministic and computed across a vast array of actors. Of course, these types of experiments and software models borrow much more from economic models than from philosophical ones, but the results are interesting because humans, to some extent, are predictable enough for these types of models to be useful. In addition, humans interacting via mediated computer technologies exhibit less trust than when interacting face-to-face making them seem more like the computer science software agents (Bos, Olson, Gergle, Olson, & Wright, 2002). Although as the technologies we use to communicate become less visible, making the interactions seem more "natural", this diminishing effect shows promise of going away. Since humans can be tested and then modeled with software (to varying degrees), this area is rich for the study of networks and untrusted interactions.

Jennifer Golbeck has done a variety of studies on networked trust and reputation models using the nascent connections making up the Semantic Web (Golbeck & Hendler, 2004a, 2004b, 2004c). She is finding that automated, local calculations of trust can be inferred and applied to systems such as email and social network inferences. This work is largely limited by the amount of online data describing people, but as the web becomes more social (Facebook, Twitter, etc.), this kind of data will become more prevalent and the models are expected to become more robust.

The idea that trust is transitive (can be sensed or passed via a trusted contact) is also prevalent in this type of work. Trust Network Analysis uses this assumption to varying degrees (Jøsang, Hayward, & Pope, 2006) and can model both positive and negative values of trust, unlike earlier work like PageRank (Brin & Page, 1998) and EigenTrust (Kamvar, Schlosser, & Garcia-Molina, 2003), and even Kleinberg's HITS algorithm (Kleinberg, 1999a). Combatting malicious use (link spam, colluding agents) is a constant struggle for the operators of social networks, but there have been gains made in engineering the social aspects to benefit the communities of users (Levien, 2004; Lauterbach, Truong, Shah, & Adamic, 2009). Post analysis of malicious use also leads to new knowledge of how to model trust networks built on real data (Gyöngyi, Garcia-Molina, & Pedersen, 2004).

2.3 Expertise

"We shape our tools, and thereafter our tools shape us." – Marshall
McLuhan (McLuhan, 1964)

Expertise is a word that is usually used to describe a sense of experience and
knowledge about a subject area. It is something that a person has acquired
and can demonstrate at will. Expertise has been studied across disciplines
and some generalizations can be applied (Ericsson, Charness, Feltovich, &
Hoffman, 2006).

Herling says expertise is defined as "displayed behavior within a specialized
domain and/or related domain in the form of consistently demonstrated actions
of an individual which are both optimally efficient in the execution and effective
in their results." This is separate from mere competence, which is minimally
efficient in the execution and effective in the results (Herling, 2000). And to
that end, it is not a mere demonstration of expertise or mastery and it is "not
an event – it's a purposeful journey" (Swanson, 2007).

It seems that expertise is something that can be generated, but it takes
time. The literature around deliberate practice and sustained effort suggest
that around 10,000 hours (roughly ten years) are necessary to accrue enough
repetitive action that things seem to come "naturally" (Gladwell, 2008). The
experts themselves do not use the word natural to describe their ability, but
non-experts in a domain certainly do.

Experts are people who can give a lay of the land quickly with a view from
above. They understand how things fit together and see a bigger picture than
those of us who may not know as much about a subject. They know where

they know things, but more importantly, they are unique in that they know where they do not know things. They are aware of their own shortcomings along the spectrum – the Rumsfeldian "known unknowns" (Rumsfeld, 2002).

Collins and Evans (2007) open their Introduction with:

> The underlying assumption of this analysis is that, other things being equal, we ought to prefer the judgments of those who "know what they are talking about." This does not mean that correct judgments are always made by those who know what they are talking about. . . . The assumption means simply that in spite of the fallibility of those who know what they are talking about, their advice is likely to be no worse, and may be better, than those who do not know what they are talking about.

Collins also does not like the relational (or labeling) view of expertise, where someone is labeled after the fact. He prefers the "realist" approach. He assumes expertise may or may not be possessed by an individual independent of whether others think they possess expertise. Contextual Authority Tagging would assume that in a sufficiently social environment over time, this becomes less true – and that the relational model holds up well enough to be considered actionable. I write more about this relative or relational aspect of expertise and truth in Section 2.3.4.

Maybury boiled all this down to five principles of expertise (Maybury, D'Amore, & House, 2002):

- *Dynamicity*: Expertise evolves over time and requires continuous awareness of changes in individual knowledge and skills.

- *Distribution*: Expertise typically resides across a set of individuals because of the complexity and breadth of technologies and missions.

- *Community*: Experts aggregate into either loosely or tightly coupled communities of expertise based on attractors such as value gained from

shared knowledge. Often experts self-organize into networks within which individuals play roles such as individual contributor, broker, or facilitator.

- *Self-Assessment*: Expertise is typically validated in the context of the work environment. Frequently peer review or assessment is a preferred validation mechanism, in part because of the scarceness of expertise.

- *Access*: Expertise is rare, expensive, and often difficult to access. Furthermore, assessments of expertise are often controlled because of privacy concerns.

2.3.1 Expert Processing

When tasked with answering what it is about experts that make them experts, most experts agree that cognitively, experts display an understanding or schema of a domain better than non-experts. Cognitively, human expertise is "characterized not by superior strategies of problem solving or a larger capacity of working memory, but larger and better selection of organized domain-specific knowledge structures (schemas) in long-term memory. Such schematic knowledge representations allow us to categorize incoming information and act in appropriate ways" (Kalyuga, 2009).

In the mid-1960s, seminal expert research showed that chess masters had extremely better recall of chess positions than those who were not chess masters (de Groot, 1965, 1966). Chess masters could remember meaningful chess placements - but did no better than non-players in remembering positions that were randomized (Chase & Simon, 1973). The masters saw real chess positions in terms of "chunks" whereas the novices saw individual pieces. The masters were seeing a different model of what was happening on the board. When

the pieces did not fit a sensical chess layout, the masters were reduced to the relatively poor performance of remembering individual pieces.

This type of cognitive modeling is well accepted now. One of the most interesting models consists of global and local architectures for processing new information (Sternberg & Frensch, 1992). When confronted with new information, experts use a global view (understanding) to first categorize the new information, but then hand off to more automatic, local processing once they "get" it. With information of a type the experts have seen before, they can easily categorize and move on with much less effort than a novice. Even when a lack of specific knowledge is available in a domain, experts use their more rigorous high-level structure and understanding of information of that type to process new incoming information or situations (Schraagen, 1993). They have a more extensive ontology in their minds that helps them to quickly assess what is new and different from what is similar to what they have seen before and probably holds little new information for them to learn from.

Similarly, experts categorize things based on existing high-level cognitive schemas, whereas novices rely on surface features of specific tasks (Schoenfeld & Herrmann, 1982). When approached with a decision or task, an expert will fit it into a larger framework of how the world works first, and then operate on what they know to be the important parts of tasks of that type. A novice will focus more on the specifics and try to solve the problem in front of them, as that is all they know. With experience and a broader familiarity, the novice begins to work more like the expert – in fact, begins to become an expert in his or her own right.

2.3.2 Deliberate Practice

Deliberate practice (Ericsson, Krampe, & Tesch-Romer, 1993) is the idea that getting better at something, obtaining success and demonstrating expertise, is a function not only of talent, but also training (Dubner & Levitt, 2006; Dubner, 2008). Excellence is accomplished mainly through the following tenets:

- Focusing on technique as opposed to outcome
- Setting specific goals
- Getting good, prompt feedback, and using it

Following these tenets, Dubner says that practice can become more about science and less about repetition. If repetition is in order, then fine, but it may not be the best thing for increasing performance.

Much of this work has been done in the area of expert performance where Ericsson et al. (2006) write in their 918-page Handbook:

> expert performers — whether in memory or surgery, ballet or computer programming — are nearly always made, not born. And yes, practice does make perfect.

This is a theme throughout Freakanomics as well (Levitt & Dubner, 2005). Levitt and Dubner use hockey players' birthdays as well as other interesting causal indicators to show that it is not just talent that makes winners winners, it is also circumstance and will. Those born earlier in the year are therefore older and bigger when the time for tryouts rolls around.

Related, but in the mainstream business literature, is Seth Godin's entrepreneurial theory of the Dip - that successful entrepreneurs and athletes

work through being average through persistence and focus. Most people burn out in the long run up to being the best – they falter in the Dip. Godin's three steps include:

- Make the world sufficiently small
- Be the best at it in that world
- Then the market cares

Getting up the hill of performance and success is about seeing the landscape and understanding that to be the best in *the world*, you have to define the scope of *the world*, and then stick to it (Godin, 2008). To take on the global market at first is probably not the smart play. Set your sights on the local market first, and be the best there. Then ramp up to the next level.

2.3.3 Expertise and the Citizen

Collins and Evans (2007) provide the basis for this next section. They clearly articulate our collective movement into "The Third Wave of Science Studies". They say that a distinction is to be made between experiences that the public has a lot of, and therefore cannot be considered specialist knowledge, and specialist expertise, in that it is notable that someone has experience and skills in an area. They feel strongly that "experts should obviously have a relatively greater input where their results are more reliable" (p135), which is to say, areas of technical expertise and social sciences. They are not talking about areas such as culture and religion.

There has been an "epistemological leveling" over the last few decades as seen in Polanyi's "Republic of Science" (Polanyi, 1962). Science has become

more approachable and knowable and "familiar," "demystified," and it is the right of everyone to be accepted into the role of scientist, assuming that norms are observed. But now is the time that we need to rebuild some of the vertical that has been leveled (Collins & Evans, 2007, 139).

The metaphor of the mountain (Collins & Pinch, 1993) is referred to directly:

> Nevertheless, to take it that the epistemological landscape is without a vertical dimension is to abandon responsibility for the world we live in. The new job of social scientists, having been so successful with the leveling, is to rebuild some structure – or, more properly, since it is obvious that there is lots of vertical structure – to understand what holds things up.

And given this mandate, the responsibility falls to the individuals of society. We need to decide who we confer power on and in what realms they should wield that power of authority. "In the absence of suitable specialist experience, the citizen can make technical judgments only through the transmutation of expertise that starts with the social expertise of ubiquitous and local discrimination – a matter of choosing who to believe rather that what to believe" (Collins & Evans, 2007, 139).

The citizen does not have to know everything or be able to prove anything on their own, they only have to have a system of credentials they can believe in and trust.

> Transmuted knowledge does not make the citizen a scientific expert capable of contributing to the question of whether it is "p" or "not-p" that is true in any particular scientific debate, but it can help the citizen make a sensible decision about whether his or her political decision should be premised on p or not-p. (Collins & Evans, 2007,

139)

The scientists themselves have a greater burden as well, through transparency and methodology. If the discussions around what science has to give to society are driven by the experts, then the "social scientists, philosophers, and other experts on expertise must [be ready to do] more than [point] to the tension between the idea of expertise and the idea of democracy." They need to guide the discussion and not just declare that a discussion needs to take place. Good policy is dependent on informed citizens and belief in a system of cognitive authority[1] where cognitive authority is necessary (when decisions need to be made by those who are expert).

The struggle between democracy (an equal vote for all) and expertise (weighted votes for those who know) is a constant balance. Getting to the point where the population feels involved but guided by knowledge instead of ideology is hard. Getting there

> means working out some way of deciding how to use expertise even when we know it is much less sure than once we thought it was, even when we know it is too early to know who the experts really are, and even when we know that it seems undemocratic to select a group of experts, however wide, to whom we grant more authority than we grant to the ordinary citizen. We must be ready to alleviate the tension between democracy and expertise by helping with the design of citizens' juries and consensus conferences: helping not just by saying "let us bring in some citizens" but by stating what kinds of citizens with what backgrounds would be best and what kind of and what length of exposure to what sort of technical

---

[1]Patrick Wilson wrote about cognitive authority as distinct from administrative authority. Cognitive authority is that which is granted to you by others because of what they think you know about. Administrative authority is that which one has because of rank or position (Wilson, 1983).

material might turn them into better representatives of the rest.
(Collins & Evans, 2007)

There will always be a mathematical distribution of expertise and knowledge within a population. There will be members with low knowledge and members with high knowledge. There will be members with amounts inbetween. And there is never an ideal amount of consensus when it is time to make a decision. The role of expertise here is to serve as shortcut for those who do not know enough to make a decision based only on the merits and details of the case at hand. These members want to and must rely on the part of the population that *can* understand the merits and details. These low knowledge members must have a mechanism by which to choose which experts to trust. It is the responsibility of the group as a whole to provide that mechanism.

2.3.4   Social Epistemology and Transparency

(spoken) Elphaba, where I'm from, we believe all sorts of things that aren't true. We call it - "history."

(sung) A man's called a traitor - or liberator
A rich man's a thief - or philanthropist
Is one a crusader - or ruthless invader?
It's all in which label
Is able to persist
There are precious few at ease
With moral ambiguities
So we act as though they don't exist

They call me "Wonderful"
So I am wonderful
In fact - it's so much who I am

It's part of my name

– Wizard, from *Wonderful* from *Wicked*

Social epistemology has relative definitions of fact and truth. These things are collectively derived from the group, from society. Interestingly, this debate comes, again, directly from "The Republic of Science" (Polanyi, 1962) having made its debut in the work of Egan and Shera (1952). The scientists learn from one another and in working together, they define what is known, what the facts are. However, the idea that facts are facts is something social epistemology has some trouble with, itself.

Truth with a capital T gives social epistemologists pause. We know that if a pencil is dropped from a table, it will fall to the floor. However, that fact is only the case within a certain framework. If we are in orbit aboard the International Space Station, the pencil will not necessarily fall anywhere. To predict future behavior requires some information about the framework in which an experiment is being conducted. The fact that the pencil will fall is only valid when the experimental framework matches up with the framework from prior experiments. And so, if it does, we can make claims on our knowledge about future events. We can make good predictions. We are pretty sure, in fact we have never seen it happen otherwise, that if we are on Earth, the pencil will fall to the floor. It is a probabilistic fact.

With more social facts of a less deterministic nature, say, the status of the former planet Pluto, the framework we are operating in defines the answer, just like before. Many textbooks still suggest that Pluto is a planet in our solar system, but it was declared a dwarf planet in 2006 (IAU, 2006). The

planet classification turns out to depend on the definition of planet itself. That definition is a social agreement between the experts who are members of the International Astronomical Union. When they changed the definition, Pluto was no longer a planet – it became a dwarf planet.

A fact is something that is established by the practice of experts. As more experts agree on a fact, we tend to think of the fact as being true. The fact itself is "socially constructed". In the same vein, a fact could be defined as a lack of contention or controversy among experts. Latour and Woolgar claim that the "reality [of a scientific entity or fact] is formed as a consequence of [the] stabilization [of a controversy]" (Latour & Woolgar, 1986, 180). Nelson (1993) even claims that the community is *required* as it is the only entity capable of knowing anything (Goldman, 2001).

In discussing the issue of justice and equality, Goldman writes "Fricker (1998) then points out that norms of credibility arise in society to pick out the class of good informants, people alleged to be competent about the truth as well as sincere. Unfortunately, societal norms of credibility tend to assign more credibility to the powerful than they deserve and to deny credibility to the powerless. The latter is a phenomenon of epistemic injustice." (Goldman, 2001)

Goldman finally comes to the conclusion that seems inevitable as soon as there is no Truth:

> It seems clear that if social epistemology is to invoke group belief and group knowledge, it should be prepared to deal with many types of groups or collectivities and many conceptions of group belief and knowledge. One size will not fit all. (Goldman, 2001)

The facts themselves may have multiple definitions. This is David Weinberger's conclusion as well – "My evidence that we're never going to agree on anything is . . . all of human history" (Weinberger, 2009). And if we are to conclude that facts are hard to agree on, assuming we can agree at all – then it must also follow that "facts are scarce" (Weinberger, 2009). When they are agreed upon or *objective*, they end discussions[2]. They help to drive out disagreement. But this is exactly the circular type of logic we end up with if we assume that facts have that kind of power. Earlier, we said facts are identified where there is a lack of controversy. Now, we are saying that facts help clear up controversy. These are a tenuous set of definitions. "Facts used to nail down arguments. Now they start them. We are in the middle of the great unnailing." (Weinberger, 2009).

Today, facts have become commodities in the sense that we assume we can find the ones we need easily and quickly when necessary. And we have turned to authority, in large part, to help make sense of our complicated, interconnected world. There are just too many things that an individual cannot possibly be expert in, and so, we defer to others. And the authority we defer to has relied on their credentials as a quality proof for what they tell us but "credentialling turns out to be a hack, based on the limitations of paper" (Weinberger, 2009). Since communication and verification have always been expensive, we took a shortcut and did the best we could; we invented certificates and credentials written on paper and we invented seals and notaries.

---

[2]I claim that mathematics is the only inductive science ($2 + 3 = 5$). All others are deductive and based on experiments (Popper, 1959) which are based on earlier established facts, measurement, and frames of reference.

But when facts require a community to define them, and the community is comprised of experts and non-experts, then deciding whom to trust becomes a new problem. The network is hyperconnected and we find ourselves dealing with multisubjectivity with regards to what it is that is known. We struggle to "triangulate towards objectivity" and find that transparency is really the only logical way we can believe what we see. Existing paper credentials do not provide the necessary transparency. We have to see the provenance of the information in front of us. We begin to demand the provenance of the facts. How strange that "transparency is the new objectivity" (Weinberger, 2009).

Transparency leads us to talk about Wikipedia.

2.3.5   Wikipedia

The Wikipedia has recently exploded on the scene and furthered the debate on power and authority and expertise quite dramatically. The implementation of a system that codifies the egalitarianism of all opinions has brought with it a resurgence of the old arguments for and against power in the hands of the practiced and knowledgeable. The Wikipedia is a "social agreement" with "no control surfaces . . . [and] completely smooth" (Pesce, 2009). In a reference to an ongoing spat between the Church of Scientology and Wikipedians, Pesce asks "What happens when the hierarchies find that their usual tools of war are entirely mismatched to their opponent?" (Pesce, 2009). Hierarchies do not interface with adhocracies, they "short out." Being transparent and dispersed, with nobody in control, and by being a series of software codes, Wikipedia lays the fundamental frameworks of our society bare and begs for further evidence

and discussion. The code demands that changes are made in public, and therefore, they are. Code is Law (Lessig, 1999).

Larry Sanger, co-founder of Wikipedia, philosopher, and staunch defender of the idea that the learned should be reserved some power of decision-making, has written about *Expertise after Wikipedia* (Sanger, 2009).

Sanger seems to assume that experts on a topic area agree with one another, but it is perfectly acceptable to think that expert opinions, on any topic, are no more consistent with one another than lay people's understanding of an issue. We can assume this to be true, especially where the topics are more soft, as he says, than the hard sciences where it is much more straightforward what are the facts and what are the agreed upon theories.

Anonymity is protected within the codes of Wikipedia. It comes with a social cost (more likely to be reverted, marked at spam, etc.) but it is a necessary means to allow information to enter the system. Subversive or unpopular opinions and people under hostile regimes all need a place to be represented and debated. Making edits and discussing potential changes with a consistent handle or user account can accrue history and reputation that anonymity cannot.

But as there is anonymity allowed (protected) within the walls of Wikipedia, the statements and claims made by authors need to be backed up with some proof. The authority of the person behind an account is not enough to push through a change, as the widely deployed annotation "citation needed" makes clear. Roles within Wikipedia do not hold authority because content gets its own authority from other places. Authority is chained, from the outside, by

design. The pillar of "No Original Research" means that authority has to be derived from other sources (Wikipedia, 2003).

Sanger asks whether this will remove the role of expert from society? The answer is no, as this is the wrong question. The wiki is a place for the dissemination of existing thought. No original research. Log the facts and frame the story here that is happening somewhere else. The reporter and the article are not the story, and they should not be.

Sanger's perspective of authority and expertise is that they are infallible and part of "truth." He addresses this while talking about the relativistic perspective, that he claims not to understand. He feels that there is truth in the world that can better be explained, conveyed, and related by someone vested in the knowledge of an area. I feel that over the long haul, the Truth, itself, changes and it is better modeled by everyone's understanding of an issue than by the experts. The scientific truths of today are dramatically different than the scientific truths from four hundred years ago. It is the burden of the expert to convey, convincingly, what their version of the truth is. It is their burden to communicate effectively why they are correct and why the masses (and other experts) should cite their opinions on how things are and what they should be.

Lanier (2006) says we are swinging too much towards the rule of the masses, a Digital Maoism. I tend to agree, but only in the sense that we currently have swung too much towards anonymity and a lack of ownership and responsibility concerning our discourse. We have comments from unnamed ogres on our newspaper sites and blogs. We lack an identity metasystem on which to hang

our names and reputations (Cameron, 2005b). On top of identity, we are provided the capabilities for reputation, respect, and real discourse.

We need a middle ground between what Lanier laments (anonymity everywhere, mob rule) and what Sanger seems to assume (experts know best and should be deferred to when they decide to intervene). We need a system that allows for anonymity, for all the reasons that anonymity is good, but also sustains a social cost for that anonymity, either by making it harder to be seen or by making verified commenters more visible. It should not be a question of blindly deferring to authority or power at the code level. The solution to this "problem" is a social one - one that requires people. Strong defaults matter, and with a default nod towards verified comments, many of the default problems with Wikipedia would melt away.

Wikipedia is not designed to be a source. A Wikipedia article should provide a shortcut compared to doing the deep digging by one's self – a first step. Wikipedia is best presented as a path towards the truth that can be found in other places. It is a list of citations and relevant work in an area. It is not, itself, the relevant work or the place where any authority lives. It *does* get its authority from outside – that is what makes it verifiable, and therefore, powerful.

Sanger cannot imagine why nobody at Wikipedia has deferred to the experts. He is driven by the vision that when presented with expert opinion, non-experts will gladly nod and say thank you and swallow it whole. This is not how our society works – and well it should not. We want cross-linking and provenance. We want skeptical thinking. We want investigative questions.

Blind followership helps no one uncover any Truth.

Cognitive authority and administrative authority; Sanger conflates the two and this is dangerous. Wikipedia articles should have no power of authority because they are Wikipedia articles. As should not the authors of those articles have any power because they are the authors. He misses the point that people are granting authority to Wikipedia because it grants authority to others, in a transparent way. The links are all clickable, allowing anyone the ability to follow a rabbit hole and decide when they, themselves, are sated. The burden is on each of us to uncover the truth that is sufficient – not on the social agreement that is Wikipedia.

2.3.6 Subjective Logic

Having spoken about facts and their lack of objectivity, we can now talk about opinions.

Measuring opinion is hard and many areas of research are active in this regard. One of the most interesting is the formal area of Subjective Logic (Jøsang, 2009). It sits between probability logic and calculus. It is a field that "models belief and confidence in uncertain circumstances." Beliefs are "rarely binary" and "usually involve some amount of uncertainty" and can be modeled with "belief distributions." The most salient feature of this field, to an outsider, is the *opinion triangle* model (Figure 2.3) that relates the variables necessary for the visual representation of uncertainty.

In binary logic, belief and disbelief always sum to one, or $b + d = 1$. In subjective logic, belief and disbelief sum to less than one, or $b + d < 1$.
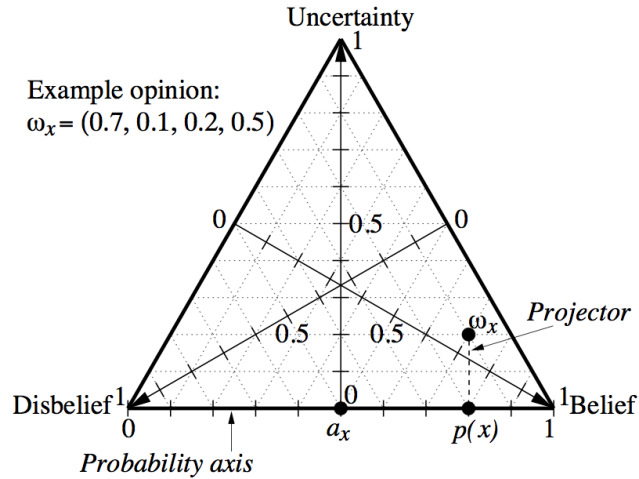
Figure 2.3: Jøsang's Opinion Triangle (Jøsang, 2009, 11). This example shows a pretty confident opinion (in the lower right corner) with belief of 0.7 and uncertainty of 0.2.

When belief and disbelief sum to less than one, this is due to some amount of uncertainty, u, represented in the final subjective logic equation $b + d + u = 1$. Together, these three elements, along with a base rate, a, combine to form an opinion on a topic, $\omega_x = (b, d, u, a)$, representing belief, disbelief, and uncertainty, with a default base rate of 1/2 (or 0.5), leaning neither with a bias towards belief or disbelief (Jøsang, 2009).

Representing opinions in this manner allows for calculations to be performed on many opinions at once and may give insight into a collective view that may otherwise have remained hidden. Jøsang says that mapping between fuzzy categories (Figure 2.4) of likelihood and certainty (Figure 2.5) could be done in a "straight-forward" manner (Jøsang, 2009, 18):

> Real-world categories would likely be similar to those found in Sherman Kent's *Words of Estimated Probability* (Kent, 1994); based on the *Admiralty Scale* as used within the UK National Intelligence

Model; or could be based on empirical results obtained from psychological experimentation.

Likelihood Categories		Absolutely Not	Very Unlikely	Unlikely	Somewhat Unlikely	Chances about even	Somewhat Likely	Likely	Very Lkely	Absolutely
Certainty Categories		**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
Completely Uncertain	**E**	9E	8E	7E	6E	5E	4E	3E	2E	1E
Very Uncertain	**D**	9D	8D	7D	6D	5D	4D	3D	2D	1D
Uncertain	**C**	9C	8C	7C	6C	5C	4C	3C	2C	1C
Slightly Uncertain	**B**	9B	8B	7B	6B	5B	4B	3B	2B	1B
Completely Certain	**A**	9A	8A	7A	6A	5A	4A	3A	2A	1A

Figure 2.4: An example set of fuzzy categories that could be mapped to Jøsang's Opinion Triangle. (Jøsang, 2009, 18)

This format presents a clean means of approaching a more mathematical model of Contextual Authority Tagging in the future. Rather than issuing beliefs that an individual knows about a topic by tagging him or her, one could additionally represent the strength of one's belief at the same time. CAT's current implementation is all or none and does not provide this level of precision.

(a) Fuzzy categories with $a = \frac{1}{3}$

(b) Fuzzy categories with $a = \frac{2}{3}$
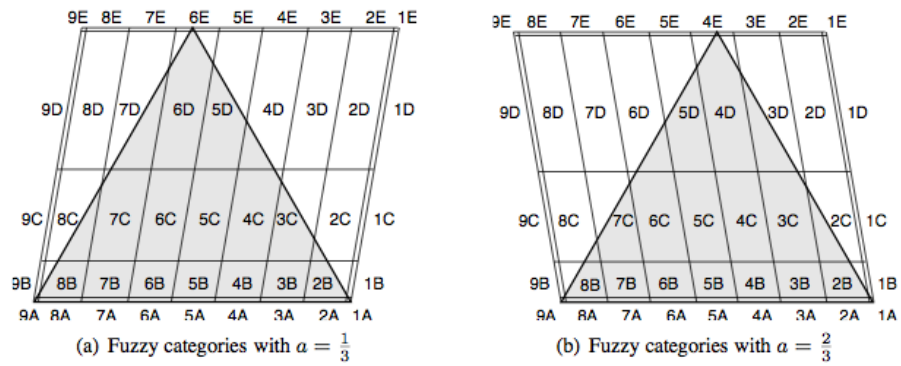
Figure 2.5: A mapping of fuzzy confidence categories to Jøsang's Opinion Triangle with two different base rate values. (Jøsang, 2009, 19)

2.4 Tagging

Tagging, or social labeling, has presented itself as a lasting effect of the Web 2.0 movement of the middle 2000s. First popularized on the social bookmarking website del.icio.us (later renamed delicious.com) by creator Joshua Schachter (Schachter, 2005), tagging has proven resilient to both detractors and traditionalists. Within a social system, tagging with simple text labels provides an inexpensive, yet rapid means of creating metadata around a set of resources (Mathes, 2004). In the case of delicious.com, these resources were URLs on the network, items that the users of the website wanted to bookmark or save for later. Instead of saving these bookmarks in simple chronological order by the datetime at which they were saved (the current best-practice of web browsers of the day), these bookmarks were saved with a set of user-added free-text labels. This proved powerful for at least three reasons: sorting and filtering through these saved bookmarks and labels became much faster and more effective, both for the user and anyone else; anyone could have a say in how something was categorized; and it encouraged the act of bookmarking to be done in public, shared for others to see and use. Other users could see (and benefit) from the actions taken by any single user. The real power came when many people saved the same link with many different words. Some of these words overlapped with other users' words and, collectively, could describe a bookmark with alarming specificity. And nobody was in charge. This collectivity without direction is what Thomas Vander Wal dubbed "folksonomy" (Vander Wal, 2007). Vander Wal also made the distinction between broad and narrow folksonomies. Narrow folksonomies are what come from a single

or few users' tagging behavior (see Flickr) whereas broad folksonomies come about through the aggregation of many users' tags (see delicious.com). He has called the tagging, or labeling, activity within an individual's Gmail account a personal folksonomy (Vander Wal, 2005).

Folksonomies are non-hierarchical, messy, non-complete, colloquial, and often sparse. But they are also nimble, esoteric, subtle, mutiplexed and often extremely less expensive to operate and maintain. And they have a much better chance, in the long run, to actually organize everything ("The only group that can categorize everything is everybody" (Shirky, 2005)). Individual users act largely in their own self-interest, with less cognitive load than traditional filing into existing categories (Sinha, 2005), and exhibit greater recall later in broad folksonomies that are highly populated (Lux, Granitzer, & Kern, 2007).

Tagging, architecturally speaking, can do everything that a hierarchical system can do. What it provides in addition, is the notion that something can be put into more than one place at a time. We have always fundamentally organized physical objects based on shelf space and available storage areas. This is because the physical object could only exist in one place at any one time, our physical space was limited and expensive, and the pre-coordination cost was necessary (we had to figure it out before we could label it). Digital information is not limited in this way. Instead of putting our documents into physical hierarchical folders, we can put many labels onto each document. Each document can exist in more than one "place" at a time. As Clay Shirky wrote emphatically, "There is no shelf!". Ontological classification works well when the domain to be organized is small, has formal categories, stable and

restricted entities, and clear, definable edges between them and when the participants are trained, coordinated, expert, authoritative users. When these things are not the case (usually, if not nearly always), we need tagging if we are going to label it (and not depend on full-text search); we need to remove the notion that we can fairly, and correctly, precoordinatedly represent the domain. Ontological classification does not work when the domain is large, unstable, unrestricted and without clear edges between entities and when the participants are amateur, uncoordinated users without the authority to declare what is true and what is not (Shirky, 2005).

When a group of people who are not trained as librarians or archivists or gatekeepers can produce a system that does many of the same things more traditional systems have been tasked with doing, but at a lesser cost and with less coordination, discussions quickly begin to form around the merits of the new system that could change everything. Thomas Friedman, when writing about the 21st century world being flat, said that the collaborative nature of online projects is "the most disruptive force of all," in how it distributes the load and the responsibility (and the gatekeeping) for deciding what is right (T. Friedman, 2005). Historically, these gatekeepers were the ones who determined what things were named and how they should be represented. David Weinberger pointed out in 2005 that "tagging repudiates one of the deepest projects our culture has undertaken over and over again: The rendering of all knowledge into a single, universal framework. The rendering has been assumed to be a process of discovery: The universe has an inner order that experts and authorities can expose. But in a networked world we know better

than ever that such an order is a myth of rationality" (Weinberger, 2005).

The world has become read/write instead of just read. Another project to take up this mantle is Fluidinfo (Jones, 2009; Baeza-Yates, Jones, & Rawlins, 2000). Fluidinfo is a storage system, a "cloud database," that only has objects and tags and users. No single user owns any object, only their own tags on those objects. The world, within Fluidinfo, is entirely read/write. Every user can have opinions and put them in public (or keep them private). The arbiters of truth become all of us. Figuring out whom to pay attention to is much harder.

2.4.1 Structure

Tagging systems are constituted by a triumvirate of user, tag, and resource (Marlow, Naaman, boyd, & Davis, 2006). Each of these three provide an essential piece of what makes these systems powerful and useful. Within a system of tagged data, a user can ordinarily pivot between looking at information about a user (what they have tagged and how they have tagged it), a tag (which users use the tag and which resources have been tagged), and the resource (which users have tagged the resource and with what tags) (Figure 2.6). Jumping among and between these views into the data, known as *pivoting*, is trivial with the hypertext upon which these systems are built.

Tagging systems in wide usage today still have not agreed upon a best practice with regards to delimiters and user interface for the addition/editing of tags. Some systems use spaces as delimiters (with no spaces in any single tag) (Schachter, 2005), some allow multi-word tags with the usage of double

Tagger

Object

Tag

@ Time t
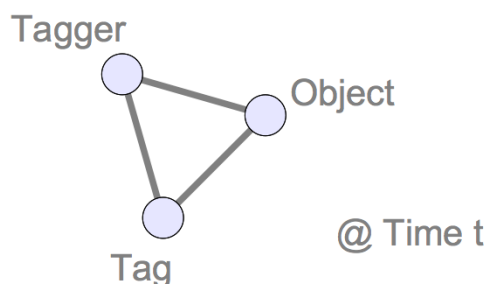
Figure 2.6: Triumvirate of User, Tag, Resource

quotation marks (seems the most awkward to use), and most are comma de-limited (which also allow for multi-word tags). There are pros and cons for each style. Flickr has an interesting procedure. They allow tags with spaces and capital letters, but store that information as the "raw" tag information and then also generate "clean" tags for each resource that they use for most internal calculations and back end processing. This allows for disambiguation between "New York", "NewYork", and "newyork" in the user interface, but allows them all to be processed by the system as "newyork" (Yahoo!, 2005).

A system that Vander Wal has found in his work to allow the most flexibility and the least confusion in a multilingual world is to have multi-word tags (allowing spaces) be entered and then managed through a series of text boxes, one for each tag. A single click can remove any existing tag (usually represented with a small graphical "x"), and there are little or no issues with confusion around commas, underscores, or spaces within a series of tags (Vander Wal, 2009).

2.4.2 Data or Metadata

In *Everything is Miscellaneous*, David Weinberger argues that data and metadata are one and the same. He feels that depending on what is interesting to you at the moment, some attributes of a thing are data and the rest are metadata. When you already know something about an object, and you are looking to find that object or a different attribute of that object, the thing you know is *metadata*. What you seek is data – it is the thing of interest. When you uncover the (previously) unknown information, that is the *data* you were looking for. Weinberger goes on to say that, while this is interesting, what is more interesting is that a different searcher could be conducting exactly the opposite search, and your metadata is his data, and vice versa (Weinberger, 2008a).

In a response to Weinberger's 2005 essay on *Tagging and Why It Matters* (Weinberger, 2005), Peterson takes issue with *Some Philosophical Problems with Folksonomy* (Peterson, 2006). She sees the two worlds of classification and folksonomy as fundamentally different and good for different things, but also feels that folksonomies confuse "cataloging structure with personal opinions" and "need to be separated". Peterson feels that the innate relativism of allowing multiple people to record conflicting annotations on a work is the single greatest concern she has with folksonomies. The author's intent should be the thing the cataloguer strives to deduce and carry out with regards to the classification of a document or work. She concludes with "Folksonomy is a scheme based on philosophical relativism, and therefore it will always include the failings of relativism. A traditional classification scheme will consistently

provide better results to information seekers."

Of course, Weinberger finds fault with this statement as well as the underlying assumption that when the classification scheme comes from above (or from the author), it is correct. He writes in 2006: "Tags are metaphysically disruptive only if one believes that (a) there is one and only one way of categorizing *The Republic*, (b) that way has to be according to Plato's intent, and (c) tags are intended to state the single, true classification of *The Republic*. If Elaine is right, then what is that true classification of *The Republic*? I don't know, I don't think Elaine knows, I don't think Plato knew, and I'm pretty sure the entire question is technically nonsensical" (Weinberger, 2006).

Later, in 2008, Weinberger agrees with himself by saying "though categorizing only by the author's intent is to me like insisting that readers only underline passages that the author considers significant" (Weinberger, 2008b). He thinks "inconsistencies in tags actually make a folksonomy useful" and are not philosophically describing the Aristotelian aboutness of an object, but rather are describing the meaning of an object to an individual searcher or reader. Who is to say that an opinion expressed by an individual, for an individual, is wrong? If proxy statements are made by others concerning the aboutness of an object based on the tags that have been inconsistently applied by searchers and readers, so be it. But that categorization is not being done by the group collectively, but rather it is an aggregated opinion, cast without collaboration, on the part of those participating in the creation of the folksonomy.

The data of what is being searched for is another man's metadata. If I find a document because the author described it a certain way, or because a fellow

searcher described it a certain way – do I care which one actually happened? Or do I only care that I found the document I was looking for in a timely manner, and that it satisfied whatever search task I was conducting at that moment? Are we all not relativists now?

2.4.3 Properties

Tag datasets that are human-generated turn out to have some fairly consistent, interesting properties very similar to many other phenomenon (Golder & Huberman, 2005). These include power law distributions of activity, social ordering effects, standard well-studied English distributions of words, and a long tail effect (a few items and a few users and a few tags show the vast majority of the activity) (see Figure 2.7).
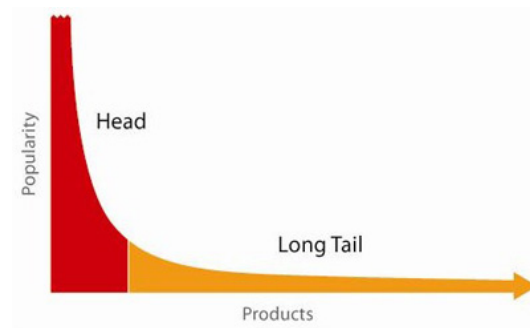
Figure 2.7: Chris Anderson's *The Long Tail* (Anderson, 2006)

One of the more interesting characteristics of a tag dataset is that when looking at a particular resource in the dataset, the usage of the tags assigned to the resource usually follow a power law as well. Some words are used a great many times in relation to most words. Quite a number of words will be used only once, or maybe twice. On the right side of a delicious.com page for a

particular URL, for example, this distribution is plain to see (See Figure 2.8).

When rotated counter-clockwise 90°, a more familiar bar chart can be drawn with the same data. This bar chart usually looks like a power law with a long tail headed down and to the right. This type of graph is extremely common in the social sciences and can be confidently placed into a family of curves called power laws. As referenced in section 2.1.5, these curves have been seen in English language usage, citation patterns, etc. Additionally, the same type of curve can be seen when data from other social media sites are collected and graphed in a similar way (Flickr.com (Dubinko et al., 2007), LibraryThing.com, and Last.fm, among others).

The word counts representing the tag usage can be plotted and graphed over time as well, as the dataset itself changes due to sustained tagging activity (Figure 2.9). This was first done at Cloudalicious in 2005 (Russell, 2006) and later confirmed and analyzed by Golder and Huberman (Golder & Huberman, 2005). The diagonal lines within such a graph are the interesting ones. Stabilization of the tagcloud is standard as time passes - the *profile* of a resource becomes familiar within a *tagspace*. What the resource is *about* is determined, at least at a particular point in time. But if a particular tag rises or falls in usage in relation to other tags, something is happening and it may indicate a need for further investigation as to the cause. The causes can be one of four types, related to the four elements of a tag cloud: 1) the *users* doing the tagging are changing (soccer moms have recently discovered this resource), 2) the *resource* has changed (actual content found at the URL, in this case), 3) the *meaning* of the words being used has changed (if we had data for the last 50

Tags	
▼ Top Tags	
ajax	1266
javascript	1089
xmlhttprequest	704
xml	610
programming	449
web	346
http	263
tutorial	201
webdev	166
html	96
xmlhttp	94
development	88
dhtml	76
reference	69
howto	59
webdesign	55
web2.0	53
code	45
dev	44
soap	43
request	40
article	37
js	34
tutorials	32
httprequest	29
rest	25
php	25
browser	25
dom	23
design	23

Figure 2.8: The Delicious.com tags for a site describing the technology that later became known as Ajax.
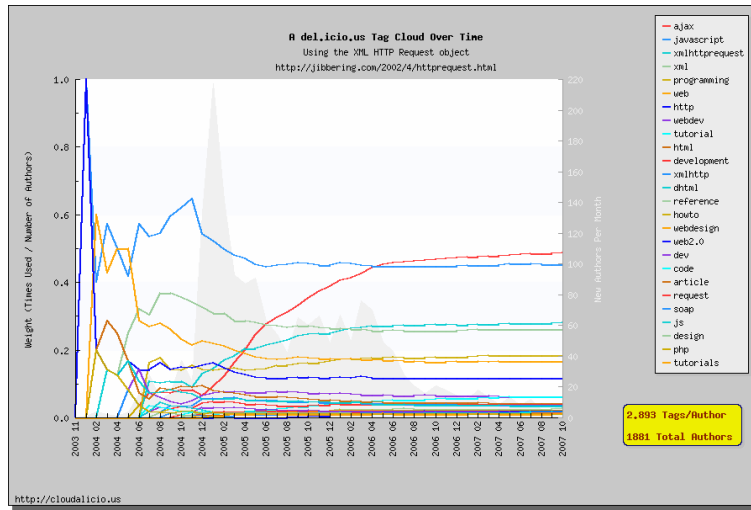
Figure 2.9: Cloudalicious - Stabilization of a tag cloud over time. Diagonals are interesting.

years, the tag "gay" and/or "happy" would have surely presented themselves as diagonal lines for certain material), or 4) *time* itself has changed the tagging practice around a particular resource as users have been influenced by earlier tagging actions (manifested in the choice of user interface into the system, or access to a system such as Cloudalicious, itself). Of course, any combination of these four types would also support a change in the tagging behavior over time (Russell, 2006).

Another type of analysis can be done when looking at time. Individual tag usage itself is not constant and the rate of this change can be thought of as a rate of decay as tags go out of favor. The tag decay of a resource, or set of resources, over time can also give insight into the "churn" of the language around an item or group of items (Russell, 2008b). If the rate of decay has been low, then the language around a resource has been more stable; there has been less volatility in the culture surrounding that particular resource (Russell,

2008a).

2.4.4 Linked Data

When the Internet was first being developed as ARPANET, the stated goal and assumption of the United States government as well as the researchers was that this network would allow and facilitate remote computing and the sharing of datasets in the course of a nuclear attack (Leiner et al., 2003). However, it quickly became apparent that messaging was the primary function of this new network and email, particularly, was the application that made the new network special. Even three decades later, as we moved towards the World Wide Web (WWW) over HTTP, data was still not the unit of interest. Rather, along with email, the webpage became the unit of record and many tools have been built around the processing, serving, and storing of webpages or documents.

As we move slowly towards a more Semantic Web, the calls have come from those who want to see the data itself be referenceable and the unit of interest (Berners-Lee, 2006). As data begins to gain first class status online, the standards around how that data should be represented and moved along the wire become critical. Recent work at the Library of Congress and at the *New York Times* have shown what is possible when existing, well documented datasets are put onto the web in standard, accessible ways. The Flickr Commons project has encouraged the Library of Congress to put over 3000 images online and asked the online community to help document and tag the information in the photographs (Oates, 2008; Raymond, 2008). The response was

remarkable and there are now over 7500 Library of Congress images, as well as many thousands from other partnerships, being publicly annotated at Flickr. The *New York Times* recently announced the availability of an effort to put all their ontological and classification data, going back over 150 years, online in the SKOS format (defined below) (Sandhaus & Larson, 2009).

One type of data that is being put online and in need of interoperability standards is tagging data. Gruber's TagOntology attempts to create a formalization around the activity of tagging and to allow description of that tagging activity at a semantic level (Gruber, 2005).

Building on Gruber's work, a few different standards have been created to work in harmony and begin to connect the vast amount of existing data sources as well as new user-created data. Figure 2.10 gives a clean representation of how four community-driven standards can be connected. Note the resemblance to the elements of a folksonomy in Figure 2.6.
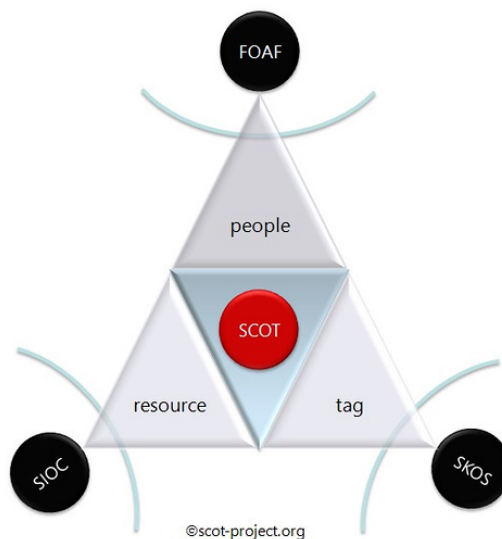
Figure 2.10: Social Semantic Cloud of Tags - SCOT Model

These four standards (SCOT, FOAF, SIOC, and SKOS) could soon become the backbone for the transport of existing and community generated tagging data and serve to populate a fledgling Semantic Web that has been struggling to get off the ground in the last decade. When data-interoperability and portability become standard features of any new tools that are built with these open standards, we should expect a blossoming of new uses and scenarios.

- SCOT - Social Semantic Cloud of Tags - Ontology for the activity of tagging (H. L. Kim & Breslin, 2008)

- FOAF - Friend of a Friend - Standard ontology for describing people and their relationships to one another (Brickley & Miller, 2000)

- SIOC - Semantically-Interlinked Online Communities - Ontology describing objects or resources in a networked environment (Bojārs & Breslin, 2005)

- SKOS - Simple Knowledge Organization System - Ontology for describing concepts, used here for tags (Miles & Bechhofer, 2009)

Contextual Authority Tagging would be producing vast amounts of time-stamped, shareable, semantically-rich data that could be published in these formats. Some very interesting social change could be just around the corner with a few loosely coupled tools that could parse, combine, and calculate based on collectively generated expertise description data. Open formats and standards are the key to new, innovative, and unforeseen uses of existing knowledge. Publishing what we know about one another in these formats could lay the groundwork for many more sophisticated later uses of the data.

2.4.5 Interface Issues

The idea of tagging people rather than resources (or URLs) is not new. Using tags for privacy and sharing control (Razavi & Iverson, 2009; Lukas, 2008) have been proposed, coded, and studied, but there has not yet been a comprehensive review of how users combine the power of tagging with the control they so desperately require over the access to their documents. The Razavi paper was a limited pilot of 10 participants in a lab setting where the participants did not have access to their own data. The Mine! project, which allows an individual to host and manage access to a repository of information via tagging, has yet to have any meaningful deployment of real users with real data (Lukas, 2008). The DiSo project aims high with goals that include hosting all of one's web data (tagging, friending, photos, writings, videos, contact information, bookmarks, locations, etc.) and sharing it out from a centralized hub - but this comes with a tremendous management overhead (DiSo Project, 2007). Facebook itself has recently required all applications to share their data through the lens of one's Friends Lists, giving users control, but again, creating a much higher cost of interaction and navigation (Facebook, 2009). But we are at the leading edge of what will very soon become normal. As how we enjoy nuanced sharing and privacy controls with the simple lowering of our voice in a restaurant, we shall soon be able to do the equivalent online. It cannot happen soon enough.

The dynamic nature of Facebook's Friends Lists are problematic as well. Razavi and Iverson talk about their OpnTag solution handling this dynamicity with ease, but they also mention at the end of their paper the potential problem of spam from others and a potential flooding from our own data. The tools

for filtering will no doubt improve, but we have not seen how they break yet in order to create better management interfaces.

If these types of interface navigations and interactions become standard for ever-smaller bits of information decision-making, we will soon be flooded with obtrusive, but menial, decisions that we will begin to ignore. Apple made fun of Microsoft's Windows Vista operating system because it began asking about every permission that a running application needed before it could execute (Cancel or Allow?) (Apple, 2007). Users obviously began to ignore or turn off these interruptions almost immediately.

Moving forward, our graphical interfaces and APIs (Application Programming Interfaces) need to provide filters and hooks so that we can dial down the amount of noise we experience when going about our computing day. Computers are moving towards being ubiquitously available and we will soon feel overloaded when having to make security decisions in addition to the ever-more-complicated tasks we are trying to complete. Humans are notoriously bad about making computer security decisions anyway, and having to make them more often, while trying to focus on other things, will only lead to more breaches and unintended consequences (Mitnick & Simon, 2001). If we can capture some of the collective wisdom about things (via tagging and social labeling), there is a much greater chance that we can keep ahead of the "bad guys" when they figure out how to present us with bogus interface decisions – they will already have been labeled as bad[3].

_____

[3]Gmail, among others, already does this with the "Report spam" button

Chapter 3

Methodology

"If you cannot measure it, you cannot improve it." – William Thomson, 1st Baron (Lord) Kelvin

3.1   Overview

I am interested in exploring the ability of a group to identify the areas of expertise of its members. Current efforts to capture this type of information almost always derive their value from either documents produced by or between group members or from asking members to talk about themselves and their own areas of expertise and knowledge. The document method is at the basis of most expert systems and knowledge management software of the past couple decades.

The self-disclosure method works, at best, when all the members tell the truth, have the best interests of the group at heart, and are thorough in their descriptions of their skillsets and knowledge. Usually, data of this kind is simply too sparse or outdated to be actionable. Members may leave out important

items from their descriptions or not participate at all. Worse, members may simply lie about their skillsets for any number of reasons.

A more robust system may be available by allowing the members to talk about each other. Holes (where things were left out) may be filled, and deception would be made more difficult because many in the group would need to give consistent, false descriptions for the collective opinion to be swayed.

If a group can (or does) know better than an individual, there should be a way to ask them. Contextual Authority Tagging may allow for the systematic gathering and evaluation of this type of information.

3.2   Delphi

The study for this dissertation used a modified version of the Delphi method. The original Delphi study was run in the 1950s and 1960s by the RAND corporation to help the US Government determine the nuclear capabilities of the Soviet Union (Helmer & Rescher, 1959; Dalkey & Helmer, 1963). They were studying the unknown military futures market by asking a variety of experts to answer a battery of questions. The answers were collated and then distributed back to the experts for additional rounds of answering the same questions - but critically, with the collective opinions of the other experts to aid their synthesis.

Rowe and Wright (1999, 354) write that, "in particular, the structure of the technique is intended to allow access to the positive attributes of interacting groups (knowledge from a variety of sources, creative synthesis, etc.), while

pre-empting their negative aspects (attributable to social, personal and political conflicts, etc.)." Over the following four decades, the Delphi method has been refined and used in many other areas besides military futures, including social science predictions (Linstone & Turoff, 1975; Rowe et al., 2005; Hsu & Sandford, 2007).

Most research has suggested that with proper preparation and consideration for expert subjects, questionnaires, and evaluation, a Delphi study can run from three to five rounds, with four being the most common number of iterations (Hsu & Sandford, 2007). Some prior Delphi studies have used post-task surveys to sample participants' reactions - from satisfaction (Van De Van & Delbecq, 1974) to confidence (Scheibe, Skutsch, & Schofer, 1975; Boje & Murnighan, 1982) to difficulty and enjoyableness (Rohrbaugh, 1979) - and I employed some of the same types of questions with CAT, especially considering the subjects were being asked to formalize their informal knowledge about one another.

A traditional Delphi study involves 1) an objective facilitator who gives "controlled feedback" in the aggregate, 2) a collection of independent experts in a domain (anonymous, to each other), and 3) a series of evaluations (iterations) designed to have the collective opinion of the experts predict the future in that particular domain (Rowe & Wright, 1999).

I modified this method to have members of a group or team define the areas of expertise for one other. This substitutes for the original formula 1) a piece of software to facilitate and aggregate free-text tags from 2) the members of the group who are anonymously tagging each other's areas of expertise in 3)

a series of rounds where cumulative tagging information is visible from prior rounds. A group of ten members, in effect, runs ten concurrent Delphis at one time – all of the participants evaluating each of the participants.

Documented criticism of the Delphi consists of lack of statistical tests, lack of demographic description of the participants, the eligibility and selection of the expert participants, the lack of explanatory quality of the responses, and the degree of anonymity of the participants (Luo & Wildemuth, 2009).

Additionally, Delphi studies need to be carefully administered to avoid the following things (Linstone & Turoff, 1975):

- overspecification of the problem statement and potential dampening of diverse perspectives

- inadequate summarization during the aggregation and synthesis stages

- lack of common interpretation by the participants of any scales being applied

- ignoring of differences of responses among participants that could be fruitful

- underestimating the amount of time and effort required to participate and administer the study

- misunderstandings between participants due to cultural or linguistic differences

Delphi has a lot to offer as a grounded, tested method to find convergence of opinion given its skeleton of domain experts, anonymity, and iteration. As Contextual Authority Tagging is being introduced to help uncover (unleash?) a collective subjective truth, the Delphi method seemed appropriate as a construct upon which to formalize this research.

3.3 Modified Delphi

I asked a team of people about their opinions, aggregated their opinions, redistributed their opinions back to the group, and then iterated the process. This process should continue either for a minimum amount of time, until their opinions "converge", or until a maximum amount of time or iterations has been met. In order to explicate the time variable, the study was performed in a series of five rounds. This was longer than I estimated to be necessary to see the tagging activity settle down, but was chosen because the amount of time required to complete an additional round was estimated to be less than one minute per participant.

As opposed to a traditional Delphi Method study, wherein the participants are selected and recognized as experts and the point of the study is to identify their collective opinion on a matter, this study used groups of people who work with one other. These group members, while not necessarily experts in any specific domain, know each other well enough to describe each others' areas of knowledge and expertise. They already grant some cognitive authority to each other in certain areas, and this study asked them to explicitly name those areas.

The irony of investigating expertise with a method originally designed to employ experts is not lost (or intentional), but I do think the approach holds up. An individual's colleagues spend more time thinking about what that individual knows more than probably anybody else, apart of the individual. They are uniquely situated to evaluate the question around the individual's areas of expertise – and therefore, I considered the colleagues the equivalent

of the selected Delphi experts. Traditionally, this type of expertise evaluation has been done either solely by the individual (via his or her résumé) or his or her boss (in a letter of recommendation or reference).

Through the anonymous aggregation and redistribution of the group members' descriptions, the areas of cognitive authority were named and quantified by the group. I employed simple keyword labeling, or tagging, as the method by which group members attributed areas of expertise to one another.

Some limitations and concerns are addressed in the following subsections.

3.3.1 Anonymity

Any concern over the **anonymity** of the participants or the attribution of the tags was reduced to the security issues around the database where the information was stored. For research purposes, I stored both the "tagger" and the "taggee", but this would not be strictly necessary if plausible deniability was of due import. Further concerns over who said what are relegated to the realm of the social – the scope of which is beyond the aim of this research. I assumed that by making these expertise tags visible and available for discussion, some stories regarding the provenance and justification of the tags were to be told. Truly secret information should remain secret, regardless of the availability of a tool or exercise like CAT – but that is an issue between those who have secrets (or privileged/private information) and those who know the secrets.

3.3.2 Selection

I also expected to hear feedback regarding the **selection** of participants of the form "friends/colleagues are not experts." I posit that they are expert, within the context in which the study is run. Some information about a participants' areas of expertise were surely beyond the purview of the other participants involved, regardless of the environment in which the study was run. That said, I limited the research to be run in professional or collegial environments where intellectual activity is the main type of interaction between participants. I explicitly avoided groups that could be construed as family, social, hobbyist, or athletic. By sticking to offices and workplaces, I expected that the types of information generated by CAT to remain largely "on topic" as that is the nature of the majority of interactions between the participants. Additional "off topic" information was generated and displayed as well, but it was expected that these tags would be limited in scope and not extend much beyond what is commonly discussed at work already; the participants will continue working together now that the study is complete. And of course, with further consideration, any "off topic" information could have been removed in later rounds by a participant who had any second guesses.

3.3.3 Misinformation

There was some concern over the possibility of **negative information** or **false claims**. These two concerns are important and deserve attention. I expected that non-normative behavior and aberrant tags would draw attention quickly. This is no different from unprofessional language being uttered or a physical

disruption in the workplace – it is quickly noticed and addressed. Negative tags were largely disincentivized by the positive phrasing of the question being asked, "What do you think this person knows about?", and "What are this person's areas of expertise?", and by the professional nature of the relationships in the groups.

3.3.4 Coverage

Another concern was that the generated tags could be argued as providing a lack of coverage and the fact that there may remain **hidden information** not captured by this technique. I agree with the possibility, but do not see it as a limiting factor. I assume that humans will always hold some information to themselves – and I encourage that. I also think that the anonymity provided by CAT allows more information than is currently being put on display to be captured and propagated. I think having total information would be a horrible thing. I also think that having a place for anonymous speech is important and that it sometimes brings potentially fascinating and useful information to the fore.

3.3.5 Statistical Rigor

Regarding the **lack of statistical measurements** and tests to determine the significance of the findings rendered by classical Delphi, I feel CAT can be claimed as immune. The nature of Delphi is that it results in a set of findings or opinions that have been deemed "convergent." The weakness of these findings can be attacked from a predictive standpoint, but as I intend for CAT to

be run continuously (if implemented beyond my dissertation research), the notion that a test was not conclusive or that there is no test is a non-issue as there are never any final "findings." The participants will take what they want from the information and use it accordingly. I see CAT being a piece of reporting/learning infrastructure that allows other tools to be built and used around it for decision making. Making the shared opinions of people visible should create more opportunities for discussion and reduce the chance for misunderstandings.

3.3.6   Loss of Control

I also expected to see some pushback from (potential) participants regarding their **not having a say** in what is being said about them and the fact that this information is being published for others to see. My counterpoint is that this is already happening, everyday, all around us. People gossip and talk amongst themselves. CAT just brings this information together, aggregates it, and shows it publicly. Damaging gossip is gossip that usually happens anonymously and behind closed doors. CAT is done in the open. Those who are good at what they do, and know their stuff, will be rewarded. Those who have not convinced their colleagues of their areas of expertise may have sparse data to show for it. Additionally, those who are well liked may be rewarded more than those who are not. This is not as much a privacy concern as it is an issue of control. CAT, I agree, definitely moves the control of defining one's areas of expertise away from the individual and towards the group (but it does not remove the voice of the individual, it just adds the voice of the group). I

also think that moving control towards the group is a good thing and something we need as we begin to live in an ever-connected, online environment where notions of identity are not as ingrained and well-understood as in our known physical world.

3.4   Lists of Tags

Conceptually, CAT employs two sets of lists - created and processed. Created lists are lists that are created by members of the group, and processed lists are lists that are the output of the process of the exercise (see Figure 3.1). They contain the same type of content (tags), but are shuffled, ordered, and aggregated differently as part of the exercise.
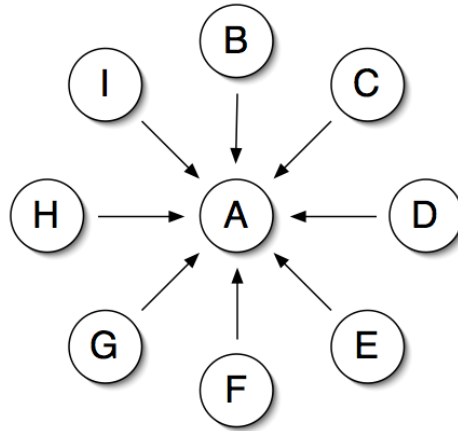
Figure 3.1: Canonical Group: Group member A is tagged by the other individuals in the group $(B..I)$ and represented as $A_B..A_I$. Collectively, the tags generated by this group about A would be represented as A^*.

The first set is the lists as they are created by a member of the group of size n. For each iteration of the exercise, each group member creates n lists.
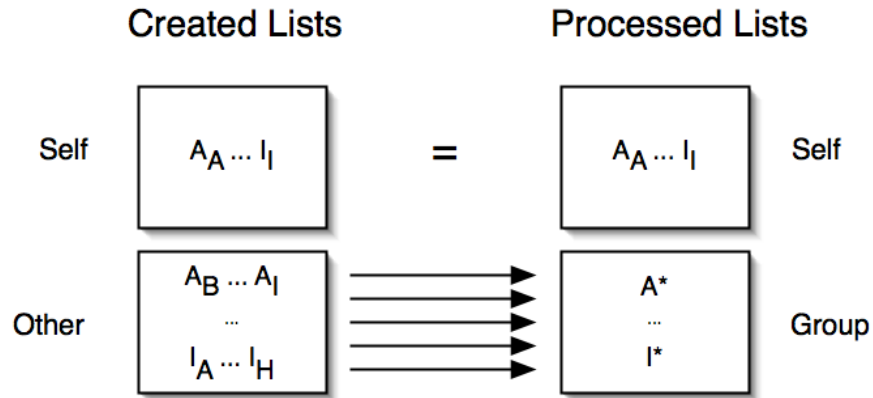
Figure 3.2: Four Lists: A group member creates lists about him/herself (A_A) and other group members ($B_A..I_A$). After processing, each member has a list about themselves (A_A) and what the group thinks he/she knows about (A^*).

They are of two types (totaling n): Self (1) and Other ($n-1$) (see Figure 3.2).

1. **Self** (A_A) : a list consisting of tags that a member uses to describe his/her own areas of expertise. There is only one Self list, per member, per iteration.

2. **Other** ($B_A..I_A$) : a set of lists created by the member to describe each of the other $n-1$ members of the group. If there are 9 total members of the group, there are 8 Other lists created, per member, per iteration.

The second set is just a reorganization by the system of the created set of lists. This set consists of the lists "about" a member, rather than "created by" a member. Each member of a group will have two processed lists describing them, for each iteration of the study.

The two processed lists include:

1. **Self** (A_A) : a list describing the individual by the individual (identical to the created Self list above), and

2. **Group** (A^*) : a weighted aggregated list where the other group members describe the individual (the Other lists combined into one)

If a group has 9 members (as in Figure 3.1), the first iteration of the exercise will generate a total of 18 processed lists, 2 for each person. If there are 5 iterations in the exercise, a total of 90 processed lists will be generated.

Within each iteration, or round, a series of four steps is followed by each member of the group. The steps include:

1. **Review** : The member is presented with the current state of the experiment from his/her perspective. His/her accumulated Self list and Group list are visible. Self and Group lists are also visible for every other member of the group. This is where most of the learning and consideration of new information presented by the software takes place. This step is used as a welcome and introduction in the first round, as there are not yet any tags to review.

2. **Self Assessment** : The member adds and removes tags to his/her current Self list of tags.

3. **Group Assessment** : The member adds and removes Other tags for each of the other members of the group.

4. **Round Complete** : The member is notified of completion of the current round.

These steps directly follow one after the other in one sitting. The spacing of the rounds was up to the groups themselves and was between two days and three weeks.

The screenshots in Figures 3.4 through 3.13 are shown as occurring in the fourth round of iterations.

3.5   Instruments and Datasets

Data collection was carried out in three major stages. Data in Stage 1 was collected via the custom tagging software and an integrated survey. Data in Stage 2 was collected through a set of semi-structured interviews. Data from Stage 3 is comprised of both human- and algorithmically-generated similarity scores. Data from all three stages is presented in Chapter 4 and then discussed in Chapter 5.

My study design is illustrated in Figure 3.3. Each group went through the process and generated tags about each member in each group.
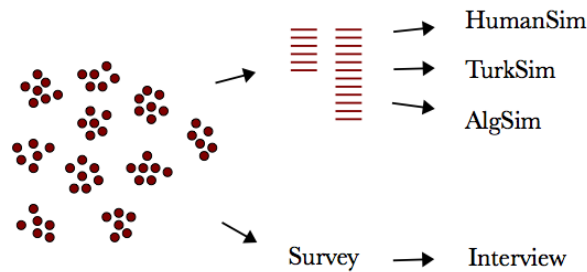
Figure 3.3: Study Design: Each group on the left is put through the Modified Delphi process. The resulting lists of words about each participants' areas of expertise are then evaluated using three different similarity algorithms. Separately, the participants complete a survey and some are later interviewed.

3.5.1 Stage 1a : CAT Software (Lists)

The custom software (seen below) generated the primary tagging dataset for this research. The participating groups each used this software through 5 rounds and populated a database of taggings (rows) consisting of a tagger, a tag, a taggee, and a timestamp. Once the groups had completed the tagging activity, the software generated 1 list of tags for each combination of group, participant, list type (self/group), and round. For an example study consisting of 6 groups with 8 participants each moving through 5 rounds, the system would generate a total of $6*8*2*5 = 480$ lists. Two of the 480 example lists can be seen in Figure 3.6, one for self and one for the group.

3.5.1.1 Screenshots
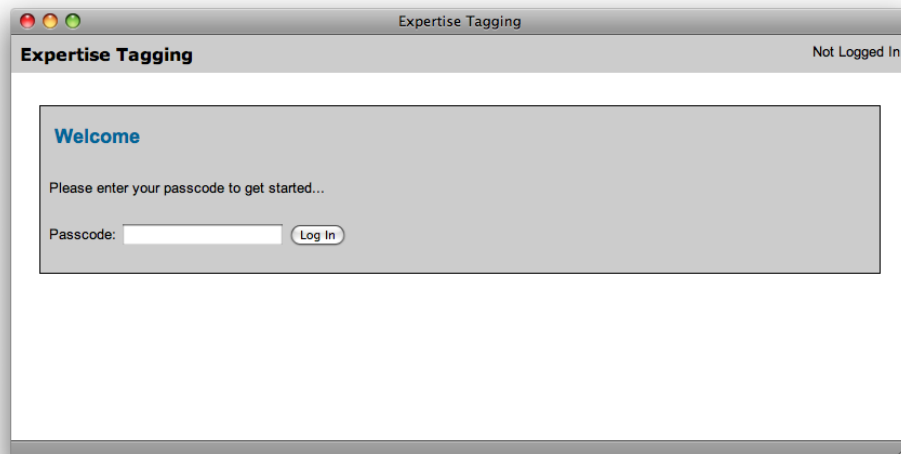
Figure 3.4: Login: Each group member used a simple passphrase to log into the system.
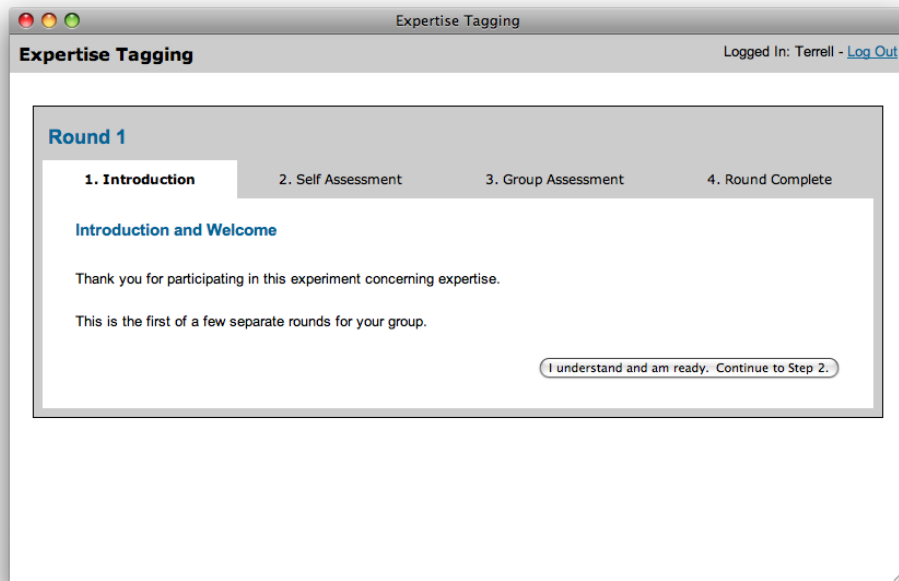
Figure 3.5: Step 1 - Introduction and Welcome: The group member is shown a welcome screen explaining the upcoming process and instructions on how to move forward. This screen is only visible in Round 1.

Expertise Tagging

Expertise Tagging   Logged In: Terrell - Log Out

Round 4

| 1. Review | 2. Self Assessment | 3. Group Assessment | 4. Round Complete |

Review - Activity through Round 3

Please review the information for each person carefully.

The first column (Self) shows how each person has tagged his/her own areas of expertise.

The second column (Group) shows how the rest of the group has collectively tagged his/her areas of expertise.
The number indicates how many different group members used the same tag.

Self   **Group**

Terrell

Self	Group
apple	**4** photography
backups	**3** apple
basketball	**3** computers
broughton	**3** tagging
claimid	**2** handball
computer_engineering	**2** internet
computer_networking	**2** library_science
computers	**2** macs
creative_commons	**2** people
digital_photography	**2** politics
expertise	**2** social_networking
handball	**2** travel
internet	(Show Single Tags)
iphone	
irods	
library_science	
mac	
mpact	
nc	
nc_state_fair	
online_identity	
open_source	
openid	
parc	
park_scholarships	
politics	
racquetball	
raleigh	
social_networking	
tagging	
unc	

(Show Group Members)

Figure 3.6: Step 1 - Review - Self: The group member is shown the aggregate listing of tags since Round 1. This includes both self tags and the aggregated tags that the group has put into the system about his/her areas of expertise.

Expertise Tagging

Kelly

Self	Group
art	**4** design
crafts	**4** pop_culture
design	**3** art
fabric	**3** journalism
fonts	**2** crafts
literature	**2** fabric
popular_culture	**2** fonts
webby_things	**2** internet
	2 park_scholarships
	2 politics
	2 quilting
	(Show Single Tags)

Simpson

Self	Group
asia	**4** china
c	**3** c
china	**3** linux
computers	**3** skateboarding
dinosaurs	**2** economics
games	**2** graphics
linear_algebra	**2** hacking
linux	**2** libertarian
lisp	**2** punk
mandarin	**2** tai_chi
mma	**2** taiwan
philosophy	**2** thailand
probability	(Show Single Tags)
reverse_engineering	
science	
taiji	
taiwan	
tea	
videogames	

Todd

Self	Group
ajax	**3** computers
bbq	**3** grilling
bears	**3** hockey
c	**3** woodworking
capitalism	**2** apex
computers	**2** babies
cooking	**2** cary
databases	**2** football
economics	**2** linux
electricity	**2** physics
finance	**2** punk

Figure 3.7: Step 1 - Review - Others: The group member is shown the aggregate listing of tags since Round 1 for each of the group members. These include both self tags and the aggregated tags that the group has put into the system about each group member's areas of expertise.

116

Figure 3.8: Step 2 - Self Assessment: The group member is asked to tag his/her own areas of expertise. The full listing of existing self tags (from prior rounds) is shown in the right column. Any existing tag can be removed by clicking on the corresponding red X.
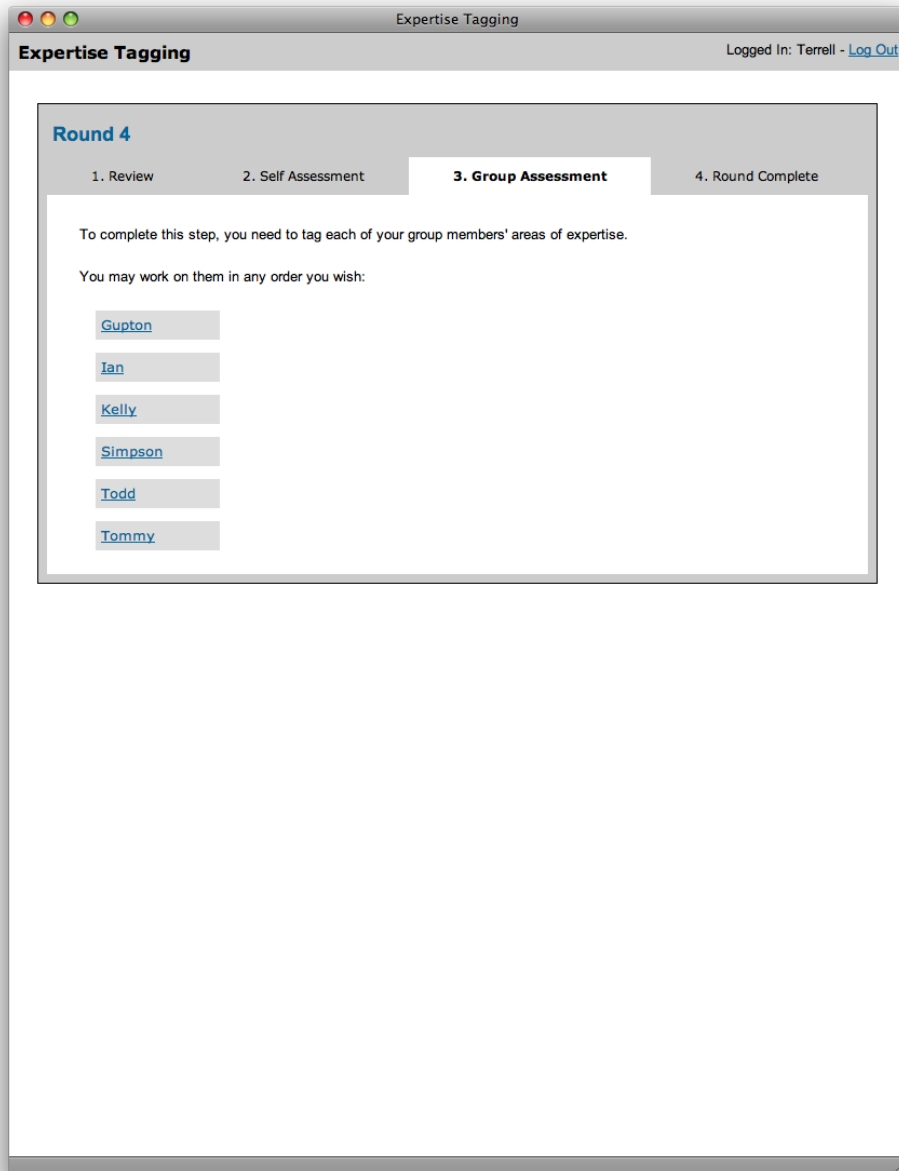
Figure 3.9: Step 3 - Group Assessment - Before: The group member is asked to tag each of his/her group members during Step 3. The group members can be tagged in any order. The group members must all be "visited" before moving to Step 4.
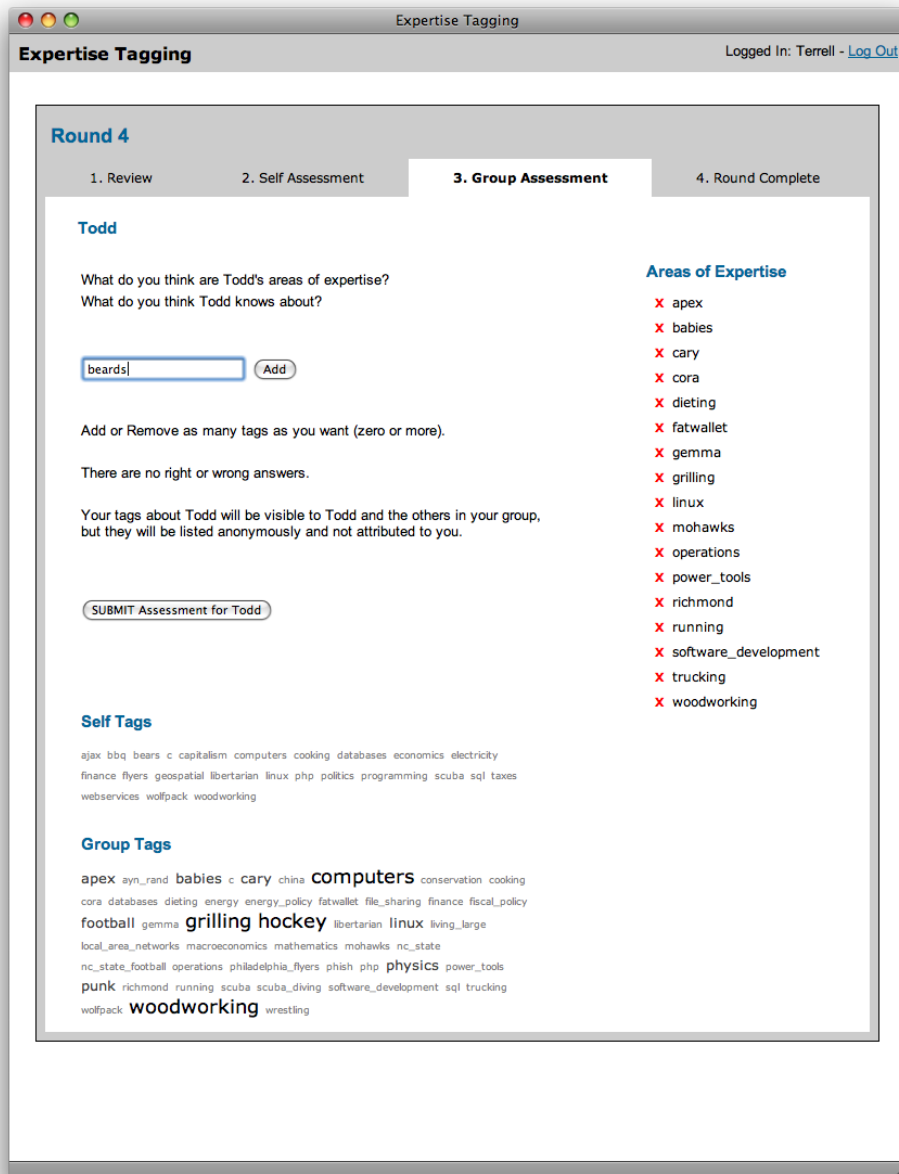
Figure 3.10: Step 3 - Group Assessment - Tagging: The logged in group member is asked to tag another group member's areas of expertise. The full listing of existing tags from the logged in group member (from prior rounds) is shown in the right column. Any existing tag can be removed by clicking on the corresponding red X.
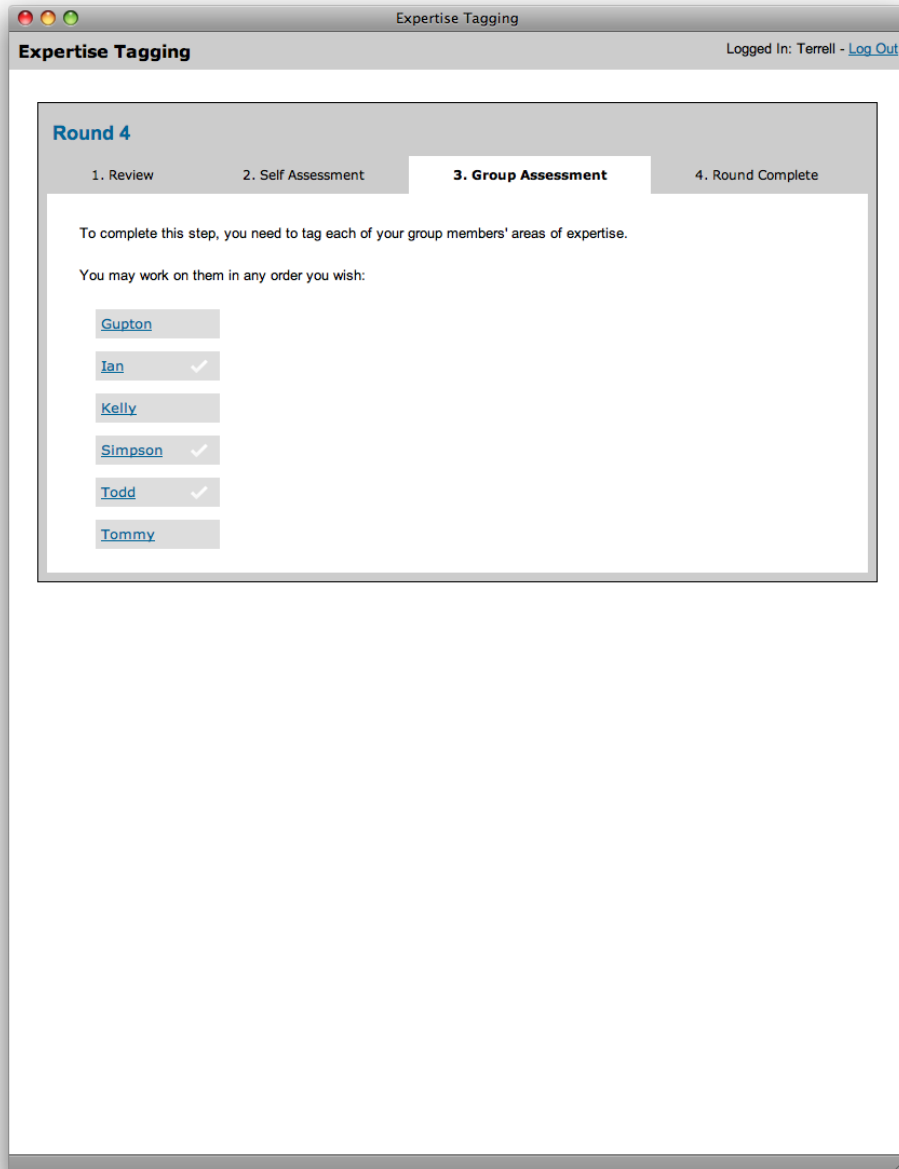
Figure 3.11: Step 3 - Group Assessment - Partial: The logged in group member must "visit" each other group member before moving to Step 4. This user has tagged 3 of 6 of his fellow group members during this round.
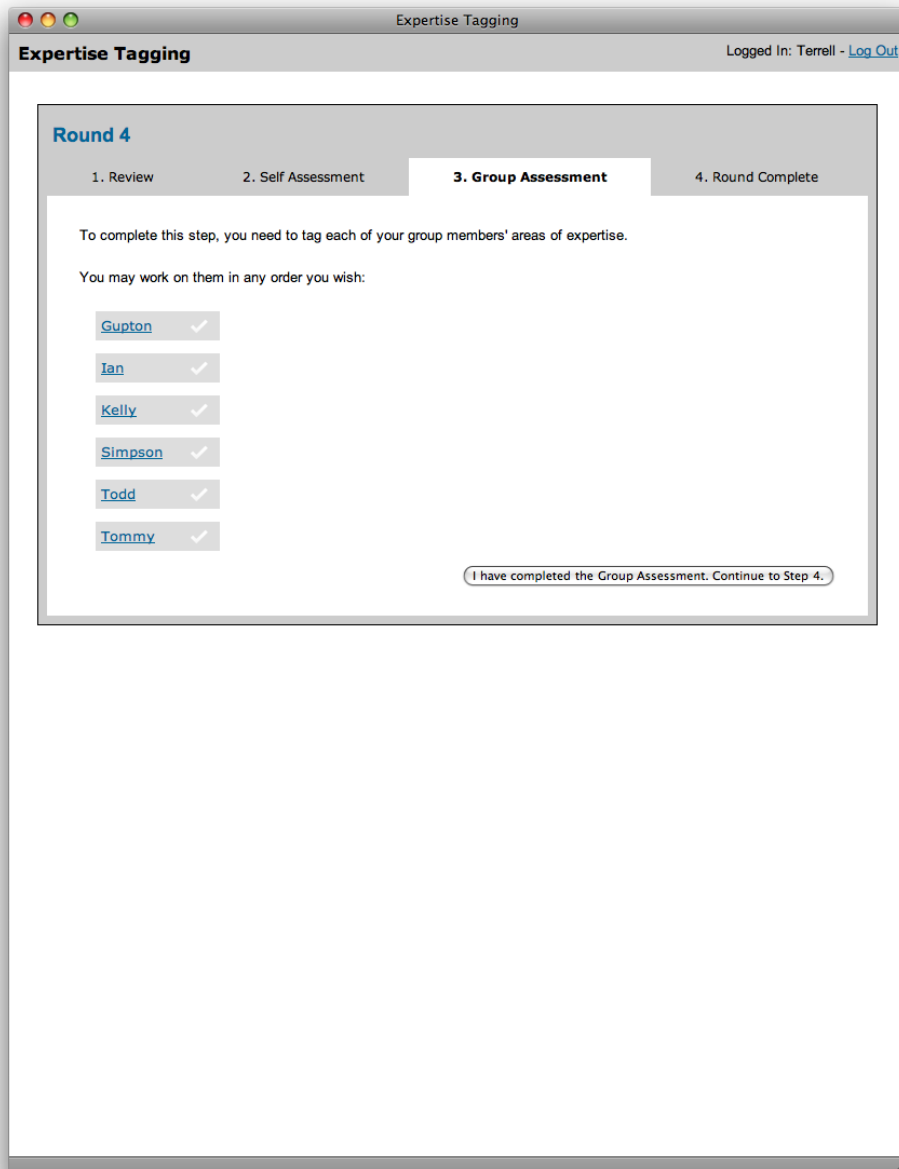
Figure 3.12: Step 3 - Group Assessment - Complete: Each fellow member has been tagged in this round. The logged in member is ready to move to Step 4.
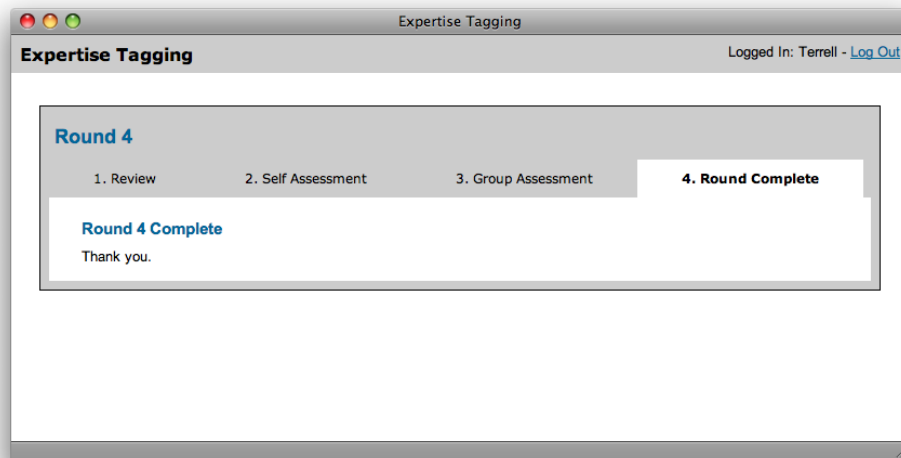
Figure 3.13: Step 4 - Round Complete: The logged in member has completed this round.

3.5.2 Stage 1b : Survey

The survey was designed primarily to answer parts of the Comfort (R2a) and Confidence (R2b) research questions. Questions marked with ∗∗ are in the form of a statement where the participant was asked to respond with a level of agreement on a seven-point Likert scale ranging from Extremely Disagree to Extremely Agree. This is in line with Likert's own scale items (Likert, 1932) and related work with the Technology Acceptance Model (TAM) (reversed in order) (Davis, 1986, 94), an instrument for measuring Adoption of Information Technology Innovation (Moore & Benbasat, 1991, 199), and the Unified Theory of Acceptance and Use of Technology (UTAUT) Model (Venkatesh, Morris, Davis, & Davis, 2003, 438).

The questions in Table 3.1 were asked in order to collect basic demographic

data about the participants.

- **Age** Under 21, 21-30, 31-40, 41-50, 51-60, Over 60
- **Gender** M, F
- **How long have you been a part of this group?** (R2a) Less than 6 months, 6-12 months, 1-3 years, 3-5 years, More than 5 years

Table 3.1: Survey: Demographic and Familiarity Items

The following sections of feedback items come from a variety of sources. The first section of items (Table 3.2) are original items designed for this study. The remaining sections of items come from prior work and have been selected because they represent validated scales designed to interrogate the potential acceptance of new technologies. Items within each particular section in Table 3.3 were averaged to produce an index. The research question addressed by each item or scale is in parentheses.

Original Items
- What was your favorite part of this exercise? Why? (R2a) Free Response
- What was your least favorite part of this exercise? Why? (R2a) Free Response
- I am familiar with my group members' areas of expertise. (R2a) **
- My group members are familiar with my areas of expertise. (R2a) **
- I am comfortable with my group's tags about my areas of expertise. (R2a) **
- I am happy with my group's tags about my areas of expertise. (R2a) **
- I would be more comfortable with my group's tags if the tags were not anonymous. (R2a) **
- My group did not list important areas of my expertise. (R2b) ** (reverse coded)
- I am confident that this system gives me good information. (R2b) **
- I am confident that this system gives me new information. (R2b, R2c) **
- I am willing to incorporate output from this system into my decision making. (R2b, R2c) **
- This was a useful exercise. (R2c) **
- This was an interesting exercise. (R2c) **

Table 3.2: Survey: Original Items

Result Demonstrability (Moore & Benbasat, 1991, 216) (R2c) ∗∗
- I would have no difficulty telling others about the results of using this system.
- I believe I could communicate to others the consequences of using this system.
- The results of using this system are apparent to me.
- I would have difficulty explaining why using this system may or may not be beneficial. (reverse coded)
Relative Advantage (Moore & Benbasat, 1991, 216) (R2c) ∗∗
- Using this system would enable me to accomplish tasks more quickly.
- Using this system would improve the quality of work I do.
- Using this system would make it easier to do my job.
- Using this system would enhance my effectiveness on the job.
- Using this system would give me greater control over my work.
Performance Expectancy (Venkatesh et al., 2003, 460) (R2c) ∗∗
- I would find this system useful in my job.
- Using this system enables me to accomplish tasks more quickly.
- Using this system increases my productivity.
Effort Expectancy (Venkatesh et al., 2003, 460) (R2c) ∗∗
- My interaction with this system would be clear and understandable.
- It would be easy for me to become skillful at using this system.
- I would find this system easy to use.
- Learning to operate this system would be easy for me.
Facilitating Conditions (Venkatesh et al., 2003, 460) (R2c) ∗∗
- I have the resources necessary to use this system.
- I have the knowledge necessary to use this system.
- This system is not compatible with other systems I use. (reverse coded)
Anxiety (Venkatesh et al., 2003, 460) (R2a) ∗∗
- I feel apprehensive about using this system.
- It scares me to think that I could lose a lot of information using this system by hitting the wrong key.
- I hesitate to use this system for fear of making mistakes I cannot correct.
- This system is somewhat intimidating to me.
Data Quality (Wang & Strong, 1996, 18) (R2b) ∗∗
- This system produced data in conformance with the actual or true values.
- This system produced data that is applicable and relevant to my job.
- This system produced data that is intelligible and clear.
- This system produced data that is easily accessible.

Table 3.3: Survey: Items from Selected Scales

3.5.3 Stage 2: Interviews

The semi-structured interviews were conducted with those who responded at the end of the survey and indicated they would be willing to be interviewed. The questions in Table 3.4 were asked over the phone and responses were recorded for transcription and content analysis. The following questions were designed to primarily answer the research questions around Usefulness (R2c).

- What was your general impression of this exercise?
- Please take me through your tags and talk about them. Value? Single term tags? Was there anything wrong or incorrect? How are the lists different? Can you characterize the difference? What's distinguishing?
- What did you learn about yourself? (R2c)
- What do you feel the group learned about you? (R2c)
- What was your favorite part of this exercise? Why? (R2a)
- What was your least favorite part of this exercise? Why? (R2a)
- How did the group feel about participating? Were they nervous? Excited? Other Emotions? Have you spoken with them since? Did you speak with them during? (R2a)
- Was the exercise a success? Has it had any effect on how the participants act towards one another? Are you satisfied with it? (R2c)
- How do you think the exercise would have been different if the tags had not been anonymous? (R2a)
- Would you recommend this type of activity to others? To partner organizations or groups? Why or why not? Who would you recommend this to, now that you have gone through it yourself? (R2c)
- Is there anything else you would like to share about this activity? About others in the group? Any interesting/unexpected tags about others? About you?

Table 3.4: Interview Questions

3.5.4 Stage 3: Similarity

The main thrust of this research is to determine whether a group and a particular member agree on a member's areas of expertise. The ratings from this dataset were used to determine this level of agreement.

The similarity datasets are designed to describe the level of similarity between the different taggings lists that are generated by the CAT software. Evaluation of subjective information (such as one's areas of expertise) must be carried out in a relative manner - as there is no objective ground truth or known yardstick against which to measure.

Three separate techniques of quantifying this similarity were used. The first used trained humans to judge the similarity of the presented sets of words. The second used untrained humans via Amazon's Mechanical Turk. The last used an existing algorithm designed to find the semantic similarity between two sentences, but without using the information encoded in sentence structure.

Graphing similarity scores against the iteration (round) shows whether the two lists (Self and Group) for a single person converge (become more similar) over time. If the similarity scores increase, then there is tendency towards convergence. If the scores do not increase, or they plateau, then there remains some difference in the lists and therefore, for a pairing of Self/Group lists, the Self and Group did not agree on that participants' areas of expertise.

This analysis can be performed for each person, then pooled and performed for each group as a whole, and then for the entire study. For the size of the datasets in this study, only group and study analyses were carried out.

3.5.4.1 Human Judged

The first dataset was generated by six trained human subjects (HumanSim). Each trained rater evaluated pairs of Self and Group lists of words and scored them on seven-point Likert scales. The interface and method used was similar to that used for the Mechanical Turk workers (see Figures 3.14 and 3.15). The trained human dataset served as the "gold standard" for the other two evaluation methods.
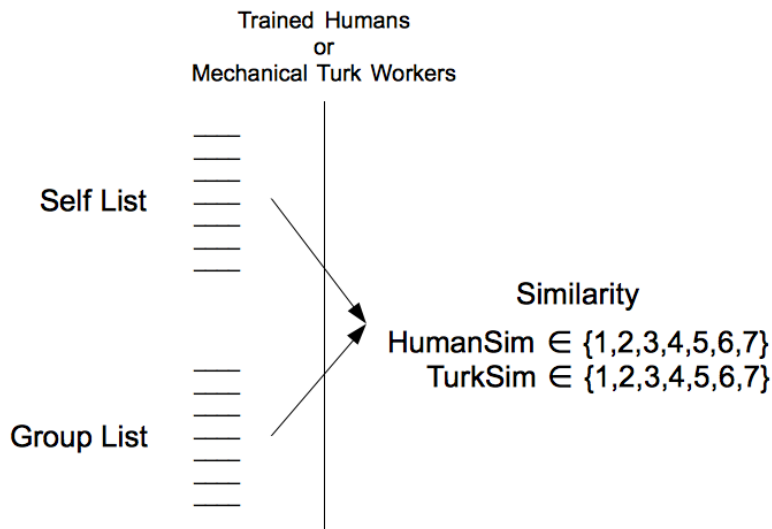
Figure 3.14: Human Similarity Rating Model – Two lists of words were presented to both trained humans and Amazon Mechanical Turk workers who were asked to rate their similarity on a seven-point Likert scale.

The second dataset was generated via Amazon's Mechanical Turk (TurkSim). Mechanical Turk is a workplace where humans have enrolled to complete simple tasks for payment. It is designed to be used for tasks that computers are ill-equipped or too expensive to solve.

HIT Preview

Similarity Rating

The following two lists of words come from different sources.

They were generated in two different ways and one list may have more words than the other.

We are interested in how similarly they describe the same concepts and ideas.

Please examine these two lists of words:

fabric	journalism
popular_culture	art
crafts	park_scholarships
literature	politics
webby_things	fabric
design	fonts
art	design
fonts	crafts
	pop_culture
	quilting
	internet

Please rate your level of agreement with the following statements:

I think these two lists describe similar concepts and ideas.

○ Extremely Disagree   ○ Strongly Disagree   ○ Disagree   ○ Neutral   ○ Agree   ○ Strongly Agree   ○ Extremely Agree

Submit

Figure 3.15: Example Similarity HIT: An example of the Human Intelligence Task (HIT) that was presented to a Mechanical Turk worker. Completion of this task was worth $0.02.

Two lists of tags were shown to a Mechanical Turk worker and the worker was asked to rate the two lists' similarity (see Figure 3.14). The similarity ratings are on seven-point Likert scales. Each pair of presented lists was evaluated five times (by different workers) to check for consistency and reliability.

The two lists that were shown are the two lists about a particular participant from a particular round. The Self list was comprised of all the self tags from that participant, randomized and shown in no particular order. The Group list was comprised of all the common tags (used at least twice by the group), unweighted, and randomized. Tags only used once by the group were removed in this similarity rating as they do not represent any type of consensus within the group.

Separate calculations were conducted by pairing the Self and Group lists when the Group list included the singly used tags, and then again when it included weighted tags.

Lastly, separate similarity scores were generated by comparing lists from random participants to generate a baseline for the turker similarity scale. Comparing "real" data lists with randomly paired lists should allow for distinguishing that the real data is qualitatively different and therefore carrying some signal.

3.5.4.2 Algorithmically Judged

The third set of similarity scores was computationally generated (see Figure 3.16) based on an algorithm (Equation 3.1) defined by Mihalcea, Corley, and Strapparava (2006). The resulting similarity scores were in the range [0..1]. The original lists of raw tags were *sense disambiguated* and then compared against one another.
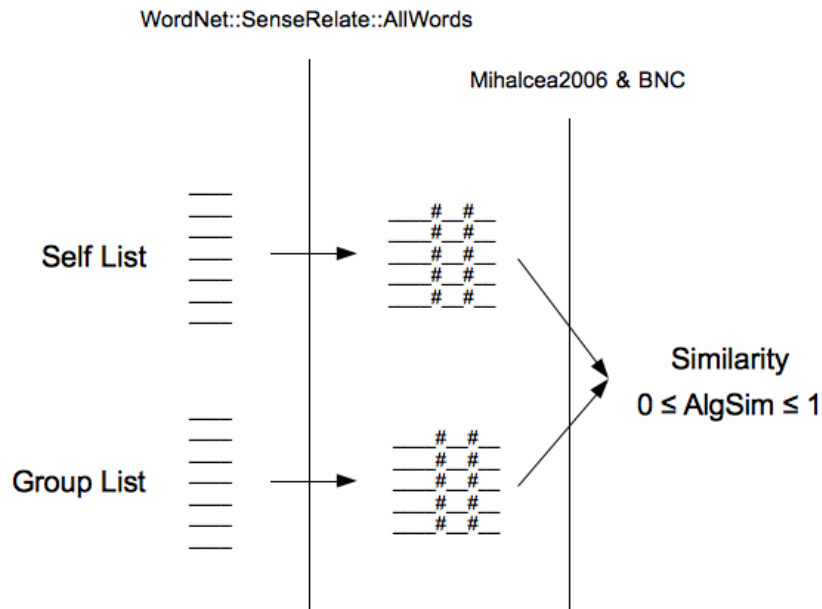
Figure 3.16: Algorithmic Similarity Rating Model – Two lists of words were 1) sense disambiguated using WordNet::SenseRelate::AllWords and then 2) compared using Mihalcea2006 (Equation 3.1) giving an AlgSim score in the range [0..1].

This method did not take into consideration the word order or "sentence" structure like more recent methods (Liu, Zhou, & Zheng, 2008). As sets of tags have no syntactic structure or order, Mihalcea was appropriate for this task.

The WordNet database was used to calculate similarity scores between two single words (Fellbaum, 1998) and accessed through the WordNet::Similarity[1] and WordNet::SenseRelate::AllWords[2] perl packages (Pedersen, Patwardhan, & Michelizzi, 2004; Pedersen & Kolhatkar, 2009). The similarity calculation is based on the "nearness" of two words in the WordNet database. Sense disambiguation uses the accompanying words in a list to detect context and assign the most probable sense to each word in the list. The inverse document frequency (idf) of a word was calculated from the 100M word sample in the British National Corpus[3] (BNC, 2007).

Tags that were not found in the WordNet database (60% of unique tags, see Table 4.4) were dropped from analysis as they could not be mathematically assessed. Tags that could not be sense disambiguated with confidence defaulted to the first numbered gloss, or definition, of the word. Tags that were sense disambiguated but then not found in the BNC were set to have an idf equal to that of the highest idf otherwise seen.

The Self list was processed as-is; each word had equal weight and all served as inputs into the model. The Group list was processed through the model in two different ways. First, all the words from the group list served as inputs, but unweighted. Second, the group list was truncated to only contain the words with a weight of two or greater. These words were then unweighted and

---

[1]http://wn-similarity.sourceforge.net/

[2]http://senserelate.sourceforge.net/

[3]http://www.natcorp.ox.ac.uk/

served as the inputs into the model.

$$AlgSim(A, B) = \frac{1}{2} \left(\frac{\sum\limits_{w \in \{A\}} (maxSim(w, B) * idf(w))}{\sum\limits_{w \in \{A\}} idf(w)} + \frac{\sum\limits_{w \in \{B\}} (maxSim(w, A) * idf(w))}{\sum\limits_{w \in \{B\}} idf(w)} \right)$$

(3.1)

Equation 3.1 took each word in set A and found the most similar word in set B (represented by $maxSim(w, B)$) and then multiplied by the information content of that word (represented by $idf(w)$). This summation was normalized across the information content of the entire list ($\sum_{w \in \{A\}} idf(w)$). After each list was compared one to the other, the similarity values were averaged for the final *AlgSim* value.

3.6   Analysis

In order to address the research questions stated at the end of Section 1.7, I conducted the analysis shown in Table 3.5.

Question	Hypothesis	Dataset(s)	Analysis
R1a - Similarity	Increasing	Lists and Similarity	HumanSim, TurkSim, and AlgSim
R1b - Convergence	Yes	Lists and Similarity	ANOVA
R2a - Comfort	Increasing	Survey and Interviews	Content Analysis
R2b - Confidence	Improved	Survey and Interviews	Content Analysis
R2c - Usefulness	–	Survey and Interviews	Content Analysis

Table 3.5: Mapping of Research Questions, Hypotheses, Data, and Analyses

Research Question 1 was addressed with the use of the Similarity datasets coming from the trained human raters, the Mechanical Turk workers, and the automatic algorithmic approach. I plotted these similarity ratings against time (Round) and expected to see the value increase. I expected the rate of change to slow over time after an initial jump in similarity ratings from Round 1 to Round 2. A repeated measures analysis of variance (ANOVA) allowed me to determine the significance of the changes over time. The different combinations of similarity ratings will allow analysis of this technique versus random pairings of lists, analysis at the group and the study level, analysis of the appropriateness of using an algorithmic technique, and analysis into the effect of displaying weighted tags versus unweighted tags.

Research Question 2 was addressed primarily with the responses to the survey and the interviews. I expected to hear a variety of perspectives on the reasons this tool created uncertainty and suspicion with regards to the participants' level of control of what they viewed as their personal information. I thought that participants would come to realize the contextualized nature of this medium of communication and that it provided a level of information that was not otherwise being captured somewhere else. With regards to confidence, CAT provides a sanity-check on what an individual thinks about someone's areas of expertise. With iteration and continued use, I thought confidence that the system was providing a unique service would increase.

3.7 Potential Ramifications

If this type of methodology and analysis can be shown to be effective in the physical world, where identity is more stable and communication channels more rich and varied, perhaps it would also work in a mediated space (an online forum, gaming, or with remote workers). When identity is more malleable and easier to manipulate, a system that can provide some infrastructure and persistence could prove very useful.

When building systems that depend on expertise tagging data for input, another potentially exciting property would be the ability to "quiet" the input from those who do not meet a certain "threshold" of knowledge in an area. If a participant (human or software agent) is not deemed knowledgeable enough on a particular topic of interest, then their input could be programmatically ignored or filtered by others. Fewer distractions lead to much higher quality discussions among those who know what they are talking about.

Additionally, on the other hand, extra voice could be given to those who *do know* what they are talking about. In an election, or key decision-making period, someone who the group deems knowledgeable in a certain domain may be routed certain questions, awarded extra votes, or have a weighted opinion counted in some other way, again, automatically or programmatically. Decisions do not have to be arrived at democratically. Most decisions in the real world are not made with equal representation.

Last, as a practical matter, organizations could use this method for deciding who to have work together when forming teams, conferences could use this method to help decide how to distribute reviewing assignments for posters

and papers, and new hires into a group or company could use this system to acclimate themselves into the culture by quickly knowing whom best to ask when they have questions.

Chapter 4

Results

This chapter will describe the results of the study and cover the study population, the tagging dataset, the types of similarity comparisons made, and the details of those different comparison techniques. Additionally, an analysis of the survey and interview data is presented.

As an overview, my study design (see Figure 4.1) consisted of 10 groups and 64 participants who generated 8773 Mechanical Turk Human Intelligence Tasks (HITs), 56 survey responses, and 15 interviews.
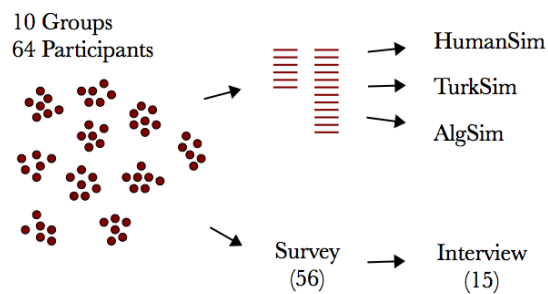
Figure 4.1: Study Design - Results: With 10 groups consisting of 64 participants, the resulting lists of words were processed through HumanSim, TurkSim, and AlgSim. Additionally, 56 surveys and 15 interviews rounded out the response data.

4.1 Study Population

A pilot study for this research was conducted using a circle of close friends and fellow PhD students. The data and the feedback from the pilot study influenced the final design of the software and the questions for the survey and subsequent interviews.

The population for this study was identified through an IRB-approved snowball sample of friends, family, and professional contacts. Each group had a primary liaison during the recruitment stage who was in contact with the researcher and gathered the names and email addresses of interested members of the liaison's group or organization. The names and emails were entered into the study software and contact from that point forward was between the researcher and the participants directly.

Nearly 200 initial recruitment emails were sent to potential liaisons and organizations. Eight negative responses came back immediately, most stating "lack of time" as the reason for not participating. Three groups responded after having a meeting of board members or team leaders and deciding it did not fit with their primary mission or direction. Fifteen liaisons brought up the proposed study with their organizations but contact was lost after three months had passed and follow-up emails were not responded to. Ten groups eventually participated in the study. The remaining contacts and groups did not respond after both initial and follow-up contact over the course of four months.

The ten groups (Table 4.1) that participated in the study consisted of 64 total participants. Two groups had five members. Four groups had six

members. Two groups had seven members and two groups had eight members. Three additional participants began the study and were tagged by their group members, but never completed the first round and subsequently dropped out.

Group	Interaction	Primary Employment	Location
family retail business	daily	yes	physical
dentist's office	daily	yes	physical
distributed software development	daily	no	virtual
distributed software development	daily	yes	virtual
museum education staff	daily	yes	physical
writer's network	not daily	no	virtual
legal non-profit	not daily	no	physical
global engineering firm	daily	yes	physical
academic faculty	daily	yes	physical
academic administrative office	daily	yes	physical

Table 4.1: Study Population by Group: 10 Groups, 64 Total Participants

The participating groups consisted of members from a family retail business, a dentist's office, two distributed software development groups, a museum education staff, a writer's network, a legal non-profit, a global engineering firm, an academic faculty group, and an academic administrative office.

Eight of the groups meet or interact on a daily basis and three are organizations that do not provide primary employment for the members. Seven of the ten groups have members who work together while physically co-present. The other three groups are dispersed and have limited or no contact in the same physical space.

The participants were not compensated for their time, but they did receive the results of the exercise (consisting of the final lists of Self and Group tags for each person in their group) at the conclusion of the study.

Of the 56 completed surveys from the 64 participants (88% follow-through

rate), there were 24 men (43%), 31 women (55%), and 1 non-response. Table 4.2 shows that the age of the participants was skewed slightly below 40 years old (55%). There was also good representation in each of the 41-50, 51-60, and Over 60 categories. The largest group of responses came from the 21-30 age group (29%). Table 4.3 shows a fairly balanced representation of group members both early and established in their membership. Over one-third of the respondents have been in their group for over 5 years (36%) and one-fourth have been in their group for less than 1 year (25%).

Age	Responses	%
21-30	16	28.6%
31-40	15	26.8%
41-50	9	16.1%
51-60	9	16.1%
Over 60	7	12.5%
Total	56	100.0%

Table 4.2: Age of Survey Respondents

Time in Group	Responses	%
Less than 6 months	7	12.5%
6-12 months	7	12.5%
1-3 years	13	23.2%
3-5 years	9	16.1%
More than 5 years	20	35.7%
Total	56	100.0%

Table 4.3: Time in Group of Survey Respondents

4.2 Tagging Dataset

The primary dataset collected through this study consisted of 10 groups, 64 participants, and over 4000 tagging events.

The study software refined the raw incoming data from the HTML text boxes presented to the participants by lowercasing the tags, substituting single and multiple spaces with a single underscore, and removing other punctuation marks and symbols in order to normalize the data as much as possible without losing semantic meaning. No automatic attempt was made to remove typos or do stemming (removing plurals, etc.).

On average, each participant labeled their own areas of expertise (Self) with 7.41 unique tags. On average, each participant's areas of expertise was labeled by their group members (Group) with 23.16 unique tags. This means that the average list of words coming back from a participant's group was three times as long as the average list of words self-assigned by that participant. I would assume this ratio would continue to rise for some time as a group's size increases.

Additionally, each participant labeled each of their group members' areas of expertise (Other) with an average of 4.43 unique tags. This is 60% of the number of Self tags per participant; participants tagged themselves more than they tagged others.

For some of the following analysis, the data was revisited and additionally cleaned by hand. Cleaning the data was attempted due to the WordNet database (used in some of the analysis) only including dictionary words. Lists of items, full sentences, and phrases are not found in the database, and therefore must be dropped, or ignored, for any automatic analysis. An average of 60% of unique last-round tags were originally dropped when running the data through WordNet (Table 4.4). Cleaning, and therefore, minimizing the loss of these data points, helped ensure the algorithm was getting as much contextual information as possible when evaluating these sets of words.

During hand-cleaning, typos were corrected, plurals were canonicalized within an individual's tags either on the singular or the plural if they did not convey distinctly different meanings. Additionally, long phrases and delimited lists were separated into their constituent ideas and words. For instance, a

	Original	Dropped	Remaining
Self	576	359 (62.33%)	217 (37.67%)
Group (all)	1688	989 (58.59%)	699 (41.41%)
Group (common)	515	255 (49.51%)	260 (50.49%)

Table 4.4: Percentage of unique last-round tags not found in the dictionary, and therefore dropped by the WordNet database. Common tags are defined to be tags occurring at least twice.

full sentence entered into the text box was cleaned to be single or double word phrases capturing the direct objects in the sentence (e.g., "Jim is an attorney with an expertise in family law. He also deals with divorce and other family matters." was reduced to the constituent tags "family law," "divorce," and "family matters" and "I like to think that I am pretty knowledgeable about our office environment and program." was reduced to "office environment" and "program"). If there was opportunity for obvious compression in the list (where a word was already in use by another tagger for that participant) then the same word was used.

After the data was hand-cleaned, on average, each participant labeled their own areas of expertise (Self) with 8.16 unique tags (a 10% increase). On average, each participant's areas of expertise was labeled by their group members (Group) with 25.16 unique tags (a 9% increase). There were 492 tagging events that were found to need cleaning (just over 17%) and this resulted in a total increase of 281 new tagging events in the database from 2834 to 3115 (a 10% overall increase).

The tagging activity was greatest at the beginning of the study (Table 4.5 and Figure 4.2). As the rounds progressed, the number of tagging events logged by the system dropped off.
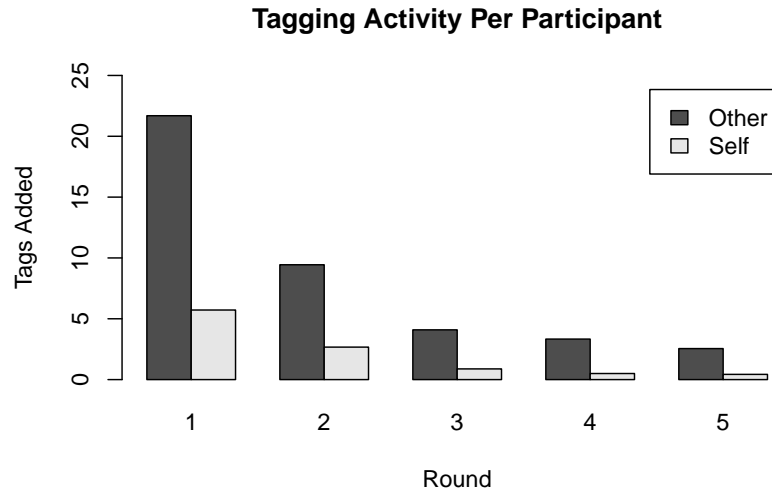
Figure 4.2: Tagging Activity Per Participant, Per Round

	1	2	3	4	5
Other	21.69	9.44	4.09	3.33	2.55
Self	5.72	2.67	0.88	0.50	0.43

Table 4.5: Tagging Activity Per Participant, Per Round

4.3 Similarity Comparisons

There were five variables taken into consideration when attempting to analyze the dataset from the participants. These represent different features that may be salient with this type of tagging data. The matrix in Table 4.6 illustrates the 2^5 possible comparison options given the five manipulated variables:

1. **Cleaned** : whether the listed words come directly from the entered data or have been hand-cleaned

2. **Random** : whether the paired lists of words (self vs. group) are from

the same participant or two random participants

3. **Group/Study** : whether the analysis is carried out at the group level or the entire study level

4. **WordNet** : whether the lists of words have been filtered through the WordNet database

5. **Weighted** : whether the lists of words are weighted or unweighted

The grey boxes highlight which combinations were analyzed for this study. Combinations without highlighting were skipped due to those combinations not making sense or being of duplicate interest. In addition, each of the highlighted combinations creates two (sets of) charts, one for all the words matching the criteria, and one for the common words (seen two or more times in the list). The values listed in the three right-most columns of the Table refer to the subsequent sections that illustrate that particular comparison.

Three techniques were developed and applied to evaluate the similarity of participant responses. They are HumanSim (Section 4.4), TurkSim (Section 4.5), and AlgSim (Section 4.6).

4.4   HumanSim

The tagging dataset was cleaned by hand for this comparison as per the guidelines in Section 4.2. The six trained raters for this set of comparisons were recruited from the researcher's peers, trained for twenty minutes each, and serve as the "gold standard" against which the later algorithms are judged.

	Cleaned	Random	Group/Study	WordNet	Weighted	HumanSim	TurkSim	AlgSim
1	-	random	group	matching	unweighted	-	-	-
2	-	random	group	matching	weighted	-	-	-
3	-	random	group	all	unweighted	-	-	-
4	-	random	group	all	weighted	-	-	-
5	-	**Random**	study	matching	unweighted	-	4.5.1	4.6.1.1
6	-	random	study	matching	weighted	-	-	-
7	-	**Random**	study	**All**	unweighted	-	4.5.2	-
8	-	**Random**	study	**All**	**Weighted**	-	4.5.3	-
9	-	-	**Group**	matching	unweighted	-	4.5.4	4.6.1.2
10	-	-	group	matching	weighted	-	-	-
11	-	-	**Group**	**All**	unweighted	-	4.5.5	-
12	-	-	**Group**	**All**	**Weighted**	-	4.5.6	-
13	-	-	**Study**	matching	unweighted	-	4.5.7	4.6.1.3
14	-	-	study	matching	weighted	-	-	-
15	-	-	**Study**	**All**	unweighted	-	4.5.8	-
16	-	-	**Study**	**All**	**Weighted**	-	4.5.9	-
17	cleaned	random	group	matching	unweighted	-	-	-
18	cleaned	random	group	matching	weighted	-	-	-
19	cleaned	random	group	all	unweighted	-	-	-
20	cleaned	random	group	all	weighted	-	-	-
21	**Cleaned**	**Random**	study	matching	unweighted	-	-	4.6.2.1
22	cleaned	random	study	matching	weighted	-	-	-
23	**Cleaned**	**Random**	study	**All**	unweighted	4.4	-	-
24	**Cleaned**	**Random**	study	**All**	**Weighted**	4.4	-	-
25	**Cleaned**	-	**Group**	matching	unweighted	-	-	4.6.2.2
26	cleaned	-	group	matching	weighted	-	-	-
27	cleaned	-	group	all	unweighted	-	-	-
28	cleaned	-	group	all	weighted	-	-	-
29	**Cleaned**	-	**Study**	matching	unweighted	-	-	4.6.2.3
30	cleaned	-	study	matching	weighted	-	-	-
31	**Cleaned**	-	**Study**	**All**	unweighted	4.4	-	-
32	**Cleaned**	-	**Study**	**All**	**Weighted**	4.4	-	-

Table 4.6: Comparison Matrix - All possible similarity comparisons. High-lighted rows are included in this analysis and can be found in the section number listed at right. Rows with no corresponding analysis were not included either because they are implausible, redundant, or did not add any obvious experimental value.

This allows for a level-setting of expectations and all-around stronger understanding for the complexity of the task at hand.

The HumanSim dataset is a sampling of all possible combinations of Self and Group lists generated by the participants. A sample was used since evaluating all combinations of Self/Group would be a monumental task (and part of the reason for using Mechanical Turk later to do so). The HumanSim dataset was created in two ways. First, the Random subset was selected by randomly selecting pairs of study participants (regardless of their group affiliation) and evaluating one participant's Self list against the other participant's Group list. This analysis determined a baseline from which to improve with the more sophisticated analyses. Second, the Study subset was selected by pairing a single participant's Self and Group lists together to be evaluated.

Each subset consisted of both weighted and unweighted samples. Each of the four cells in the 2x2 design consisted of 30 sampled pairs for each of the five rounds both when all the tags were used and when only the common tags were used (used more than once). Each pairing was evaluated by two independent raters. Each of the four cells in Table 4.7 and Figure 4.3 therefore consisted of nearly 600 evaluations ($30 * 5 * 2 * 2$), with some pairs being dropped due to duplication within the sample.

	Random	Study	Total
Unweighted	588	584	1172
Weighted	586	590	1176
Total	1174	1174	2348

Table 4.7: HumanSim: 2x2 Design. A total of 2348 comparisons were made by the six human raters.

The HumanSim evaluations were executed and managed through a separate custom web application and was very similar in form and function to the interface presented to the Mechanical Turk workers (Figure 3.15). The same 7-point Likert scale was used and each pair was evaluated in an average of 15 seconds. Due to little apparent analytic value at the edges of the scale, ratings of 1, 2, and 3 were collapsed in this reporting to "Low" similarity and ratings of 5, 6, and 7 were collapsed to "High" similarity. Ratings of 4 remained "Neutral".
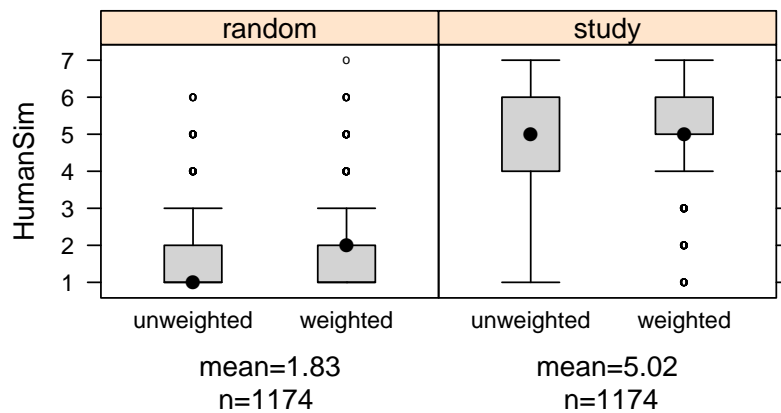

Figure 4.3: HumanSim

A two-way analysis of variance shows that both independent variables were significant and had no significant interaction effects. The main effect between the random design and study design was significant with a p-value of 0.000. The main effect for weightedness was significant at the 0.01 level with a p-value of 0.0013.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
random.study	1	5995.53	5995.53	3656.91	0.0000
weighted.unweighted	1	16.95	16.95	10.34	0.0013
random.study:weighted.unweighted	1	0.91	0.91	0.55	0.4566
Residuals	2344	3843.01	1.64		

Table 4.8: HumanSim: Two-way ANOVA

The Random subset of list pairings were rated to have a definitive dissimilarity with a "Low" average rating of 1.83 (Figure 4.3). There was no statistically significant difference (p=0.054) between the weighted and unweighted lists in this subset (Table 4.9). This subset served as the null hypothesis and the baseline from which all other rating comparisons were judged. Trained humans can determine that randomly paired sets of words are not similar.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
weighted.unweighted	1	5.01	5.01	3.73	0.0537
Residuals	1172	1573.58	1.34		

Table 4.9: HumanSim - Random: ANOVA, comparing weighted vs. unweighted

The Study subset of list pairings were rated to have a definitive similarity with a "High" average rating of 5.02 (Figure 4.3). There was a marked difference between the weighted and unweighted lists in this subset (statistically significant with p=0.01, Table 4.10). Lists that had weights associated with the listed terms were evaluated as more similar to one another, mostly by removing the ambiguity on the low end of the scale. Trained humans determined that labels attributed to one's areas of expertise by one's peers are similar to the labels given by someone about their own areas of expertise.

147

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
weighted.unweighted	1	12.86	12.86	6.64	0.0101
Residuals	1172	2269.43	1.94		

Table 4.10: HumanSim - Study: ANOVA, comparing weighted vs. unweighted

In addition to the categorical and means analysis, the percentage of agreement between the raters is important. Figure 4.4 and Table 4.11 illustrate the distribution of agreement between the trained human raters.
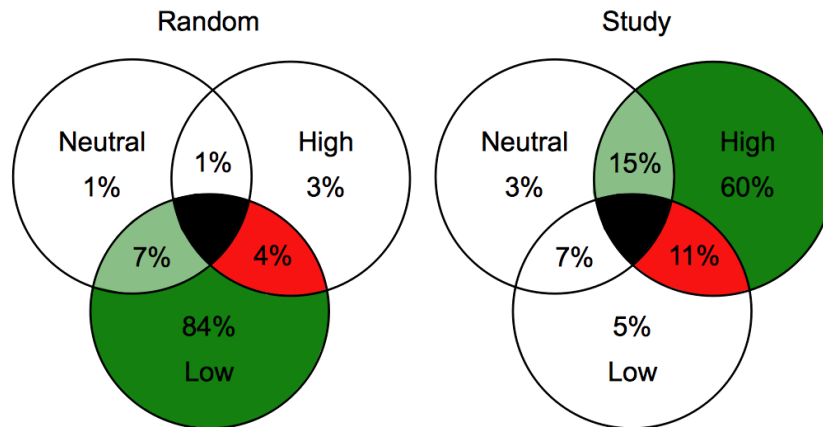
Figure 4.4: HumanSim Agreement – Only 60% agreement between humans when looking at Self/Group lists about the same person.

There were a few interesting results that follow from this analysis. First, interrater agreement for the Random subset was very strong with both raters agreeing that the lists were dissimilar 84% of the time. This rises to nearly 92% of the time when including the scenario where one rater thought the lists were neither similar or dissimilar. When looking at the Study subset, the agreement was not as strong. Surprisingly, the raters only agreed that the lists were similar to one another 60% of the time. When including the times

when one rater was not as sure, this rises to almost 75%. While this number is not as high as expected, it shows that agreeing on semantic similarity is a hard problem and not consistent in the minds of different individuals.

Specifically, the raters' feedback included that it was especially hard to decide whether the lists were similar or not when one list was a complete subset of the other or when the two lists of words were dramatically different in length. Sometimes, the raters felt there was some "missing information" from one list or the other, some part of a puzzle they were not being given. Other times, they reported feeling confident in their assessments "except for one word" which would throw them off, a kind of "noise."

Similarly, the disagreement rate (when one rater evaluated High and the other evaluated Low) exhibited a complementary pattern. The Random subset of similarity ratings showed High/Low disagreement under 4% of the time while the Study subset presented a more cloudy 11%.

These figures illustrate that, with minimal training, the human raters were able to largely evaluate the same inputs with the same outputs and distinguish random comparisons from real data.

	Random		Study	
Both High	15	2.6%	354	60.3%
Neutral/High	8	1.4%	85	14.5%
Both Neutral	4	0.7%	17	2.9%
Neutral/Low	43	7.4%	39	6.6%
Both Low	490	84.3%	27	4.6%
High/Low (disagreement)	21	3.6%	65	11.1%
Total	581	100.0%	587	100.0%

Table 4.11: HumanSim Agreement – Total unique HumanSim evaluations: 1168

The human raters, themselves, were fairly consistent in their volume of disagreement. Table 4.12 shows that each of the six raters in the Study subset had an "error" rate of only 3-8%.

Rater	High/Low Ratings	Percentage
1	26 of 392	6.6%
2	19 of 392	4.8%
3	29 of 390	7.4%
4	20 of 392	5.1%
5	22 of 390	5.6%
6	14 of 392	3.6%

Table 4.12: HumanSim Rater High/Low Disagreement in the Study subset

The takeaway from this HumanSim dataset seems to be that if humans are agreeing that this type of input presented in this way is similar only 60% of the time, then any algorithm we build to duplicate human analysis will have a hard time deciding what is similar as well.

4.5 TurkSim

The second of three techniques used to evaluate the similarity of the study data was via Amazon's Mechanical Turk. This section describes the results of this set of evaluations with both Random pairings as well as analysis at both the Group and the entire Study level.

In its entirety, this study cost \$219.33 and was conducted via a series of 8773 Human Intelligence Tasks (HITs) within the Mechanical Turk system.

4.5.1 Random

The random pairing analysis in this section is similar to what was done in the previous section (4.4) with trained human raters. Each pair of lists shown to the Turkers is a combination of a Self list from one participant with the Group list from a dif-

from Table 4.6, Line 5
- not cleaned
- random
- entire study
- in WordNet
- unweighted

ferent random participant. The similarity of these two lists was expected to be low and, therefore, this set of scenarios serves as the null hypothesis for the Mechanical Turk (TurkSim) evaluation technique. Results that do better than Random are exhibiting some effect of the study.

The subset of pairings here represent random participants, but only using the words that matched with words known to the WordNet database. Words that were not in the WordNet database were dropped in this analysis and not listed when evaluated by the Turkers so that this analysis is directly comparable to the equivalent AlgSim evaluation later (which cannot, by definition, evaluate words not in WordNet as they have no WordNet definitions or weight). Additionally, these pairs were not weighted (duplicate words were not designated as such), so every word included in the lists was only included once per list. The words were listed in alphabetical order.

Figure 4.5 shows that the ratings for the "TurkSim - Random" subset were situated clearly on the Low similarity end of the scale. The mean similarity score when "all" the words were listed was 3.0 and 2.8 when only the words occurring twice or more were listed ("common"). The median score for both
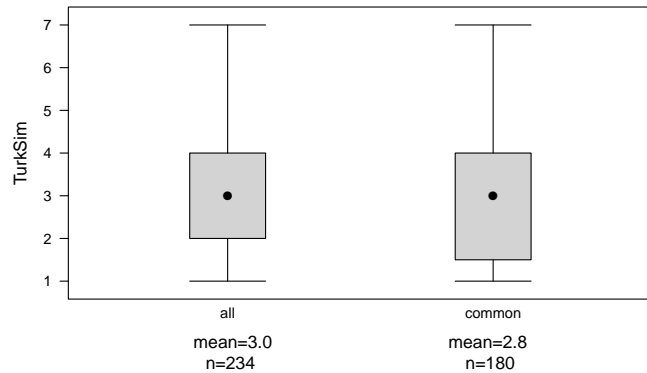
Figure 4.5: TurkSim - Random

was 3. The minimum for both was 1 and the maximum for both was 7.

4.5.2 Random - All Words

The subset of pairings here represent random participants, using all the words used by the participants, not just those found in (matching) the WordNet database. These pairs were not weighted, so every word included in the lists was only included

from Table 4.6, Line 7
- not cleaned
- random
- entire study
- all words
- unweighted

once per list. The words were listed in alphabetical order.

Figure 4.6 shows that the ratings for the "TurkSim - Random - All Words" subset were also clearly on the Low similarity end of the scale. The mean similarity score when "all" the words were listed was 3.5 and 3.3 when only the words occurring twice or more were listed ("common"). The median score for
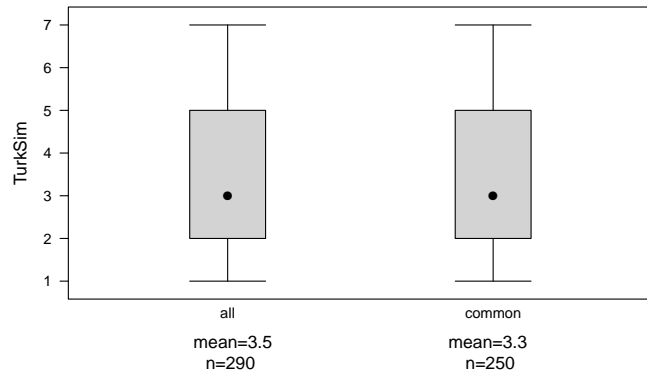
Figure 4.6: TurkSim - Random - All Words

both was 3. The minimum for both was 1 and the maximum for both was 7.

4.5.3 Random - All Words - Weighted

The subset of pairings here represent random
participants, using all the words used by the par-
ticipants, not just those found in (matching) the
WordNet database. These pairs were weighted, so
the words included in the lists were annotated with

from Table 4.6, Line 8
- not cleaned
- random
- entire study
- all words
- weighted

a "score" designating how many times that word was used in the list. The

words were listed by weight, with the heaviest words first. Words with identi-

cal weights were listed in alphabetical order.

Figure 4.7 shows that the ratings for the "TurkSim - Random - All Words -

Weighted" subset were also clearly on the Low similarity end of the scale. The

mean similarity score when "all" the words were listed was 3.3 and 3.1 when

only the words occurring twice or more were listed ("common"). The median

score for both was 3. The minimum for both was 1 and the maximum for both was 7.
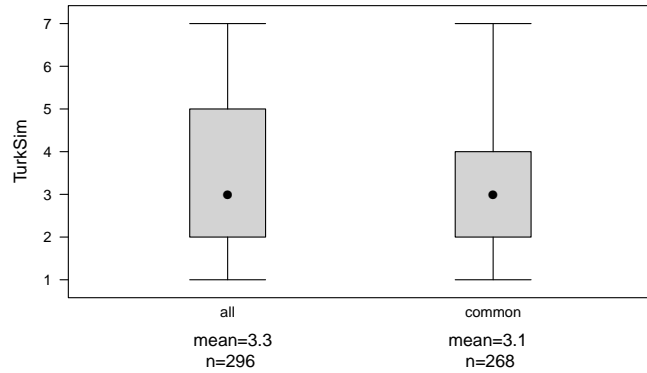
Figure 4.7: TurkSim - Random - All Words - Weighted

All said, the set of Random pairings suggest that untrained Mechanical Turk workers can determine that randomly paired sets of words are not similar.

4.5.4   Group

These results are presented by round and can be interpreted from left to right with respect to time. Time elapsed between rounds is not uniform but the axis still serves as a proxy for the passage of time and increasing participant familiarity and usage of the study tool.

from Table 4.6, Line 9
- not cleaned
- not random
- by group
- in WordNet
- unweighted

There are missing boxplots in Figure 4.8 for the last two rounds of Groups 3 and 4, and for the last round for Group 6, as these groups only completed 3 and 4 rounds, respectively. These will be missing for all Group analysis (this

and the next two subsections). The data from these rounds were not sent to the Turkers, as the additional paid results would simply duplicate the previous round that added new data.

Additionally, there are missing boxplots for Groups 1 and 6, Round 1, and Group 10, all rounds, since there were no terms that were known to WordNet that also were mentioned by multiple members of the group. As such, there was nothing to send to the Turkers to compare.

In general, there was a slight upward trend in similarity ratings over time. The Turkers generally agreed that the lists of words provided by the participants were more similar than dissimilar (ratings of 5 and above). The two to three groups with lower similarity (ratings of 3 and below) were using more specific and technical language to define their areas of expertise. This language may not have been as accessible to or known by the Turkers and rated accordingly.

all				
g1	g2	g3	g4	g5
70	100	100	60	145
g6	g7	g8	g9	g10
90	135	160	120	100
common				
g1	g2	g3	g4	g5
45	80	95	30	130
g6	g7	g8	g9	g10
40	135	160	120	0

Table 4.13: TurkSim - Group (Totals): These are the number of Mechanical Turk HITs evaluated for each group and shown in Figure 4.8. The totals vary due to the size of the group memberships and the specifics of the language used by the participants. Only words found in WordNet were used in this analysis.
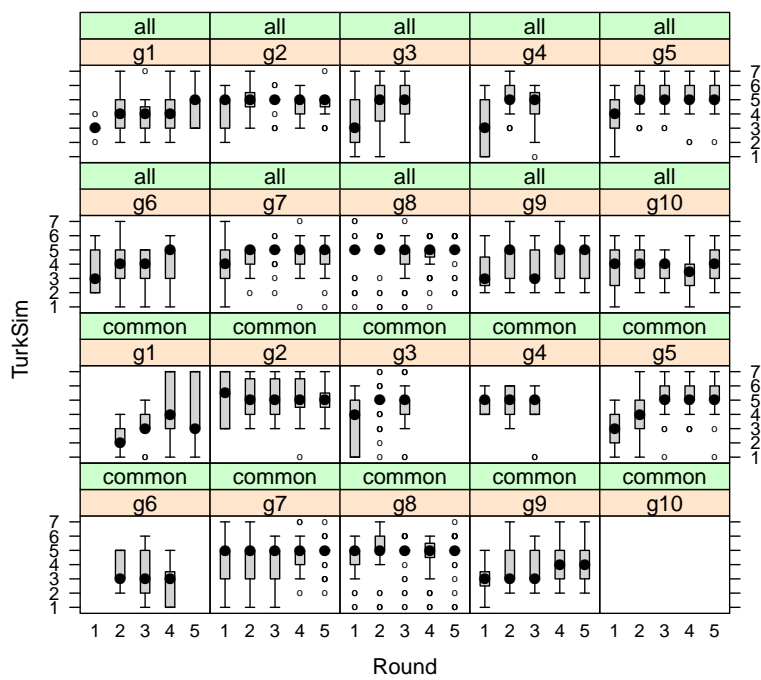
Figure 4.8: TurkSim - Group

4.5.5 Group - All Words

The ratings in this section are from the Turk-
ers when presented with the Self and Group lists
of participants.   These lists were presented un-
weighted and included all the words given by the
participants (no words were dropped due to lack of
inclusion in the WordNet database).

from Table 4.6, Line 11
- not cleaned
- not random
- by group
- all words
- unweighted

Again, Groups 3, 4, and 6 in Figure 4.9 are missing their final round(s) due
to those groups not completing the entire study. Group 10 is missing data due
to no common terms (used by more than one group member) being present by
Round 1.

These ratings are similar to the WordNet results in the last section, but
noticeably higher for the 2-3 groups that had earlier exhibited ratings in the
lower end of the similarity range. Due to more words being present for the
Turkers to evaluate, they rated the lists higher, in general.

4.5.6 Group - All Words - Weighted

The ratings in this section are from the Turk-
ers when presented with the Self and Group lists
of participants.   These lists were presented in a
weighted fashion (words were shown with a corre-
sponding number representing how many times it
was present in the list) and included all the words given by the participants

from Table 4.6, Line 12
- not cleaned
- not random
- by group
- all words
- weighted
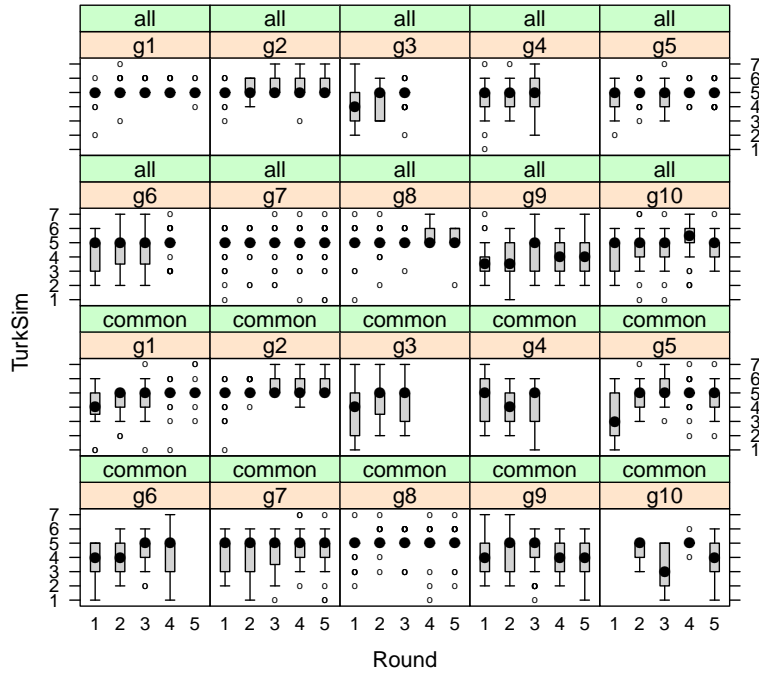
Figure 4.9: TurkSim - Group - All Words

all				
g1	g2	g3	g4	g5
125	150	105	75	150
g6	g7	g8	g9	g10
150	195	175	150	145
common				
g1	g2	g3	g4	g5
115	150	105	75	150
g6	g7	g8	g9	g10
120	190	175	150	20

Table 4.14: TurkSim - Group - All Words (Totals): These are the number of Mechanical Turk HITs evaluated for each group and shown in Figure 4.9. The totals vary due to the size of the group memberships and the specifics of the language used by the participants.

(no words were dropped due to lack of inclusion in the WordNet database).

Again, Groups 3, 4, and 6 in Figure 4.10 are missing their final round(s) due to those groups not completing the entire study. Group 10 is missing data due to no common terms (used by more than one group member) being present by Round 1.

By visual inspection, these ratings are extremely similar to the ratings from the last (unweighted) section. There is little change when introducing weighted listings to the Turkers. The ratings were generally scores of 5.
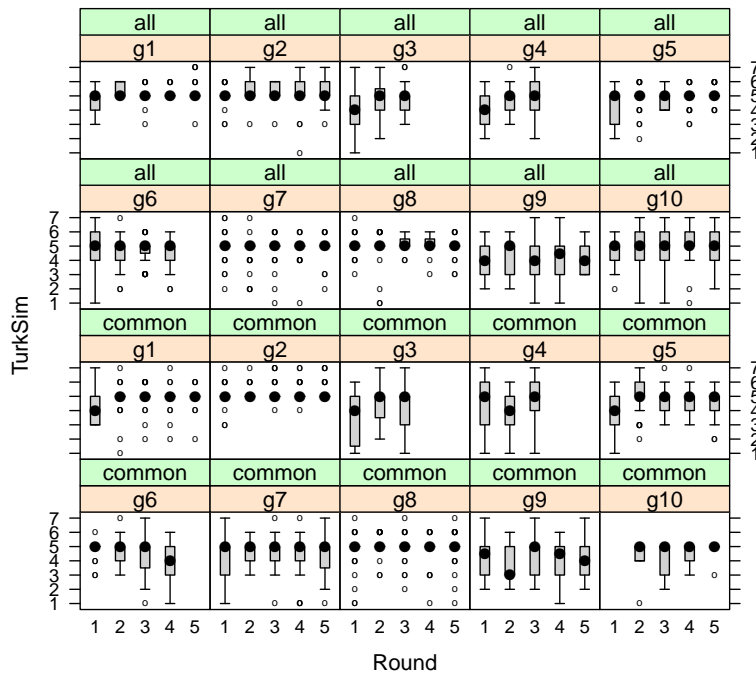
Figure 4.10: TurkSim - Group - All Words - Weighted

		all		
g1	g2	g3	g4	g5
125	150	105	75	150
g6	g7	g8	g9	g10
150	195	175	150	145
		common		
g1	g2	g3	g4	g5
115	150	105	75	150
g6	g7	g8	g9	g10
120	190	175	150	20

Table 4.15: TurkSim - Group - All Words - Weighted (Totals): These are the number of Mechanical Turk HITs evaluated for each group and shown in Figure 4.10. The totals vary due to the size of the group memberships and the specifics of the language used by the participants. This table has identical values as Table 4.14 because the same number of evaluations were made. These evaluations were made when the tags were weighted.

4.5.7   Study

The ratings in this section represent the composite of all the group-specific ratings displayed in Section 4.5.4. As seen in Figure 4.11, when the data was filtered through WordNet, there was no confidence by the Turkers that the words were similar in Round 1 (average rating of 4, or Neutral). The similarity ratings increased slightly by Round 2 and onward as the rounds continued. Having only the common words shown to the Turkers decreased their ability, at the margins, to discern between the lists, as the range of ratings were wider when they

from Table 4.6, Line 13
- not cleaned
- not random
- entire study
- in WordNet
- unweighted
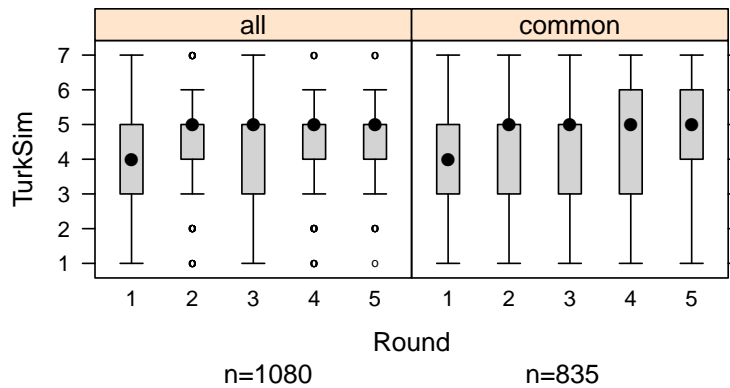
Figure 4.11: TurkSim - Study

had less information to evaluate.

4.5.8 Study - All Words

The ratings in this section represent the composite of all the group-specific ratings displayed in Section 4.5.5. Since this data was not filtered through WordNet, there was more information for the Turkers to use in deciding their similarity scores.

from Table 4.6, Line 15
- not cleaned
- not random
- entire study
- all words
- unweighted

The ratings in Figure 4.12 were "Similar (5)" in Round 1 and remained there for the rest of the rounds, but the range tightened noticeably as the rounds continued. The ranges collapsed more quickly (Round 2 vs. Round 5) when all the words were shown, rather than only the words that appeared in

161
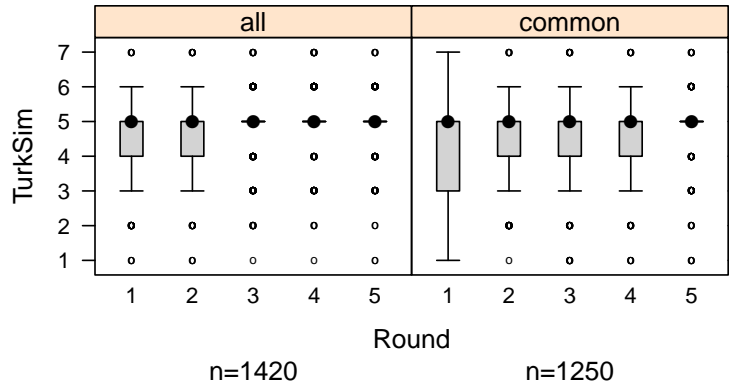
Figure 4.12: TurkSim - Study - All Words

the lists multiple times.

4.5.9 Study - All Words - Weighted

The ratings in this section (see Figure 4.13) represent the composite of all the group-specific ratings displayed in Section 4.5.6. Having the data be presented in a weighted fashion to the Turkers seemed to have little to no effect on their similarity

from Table 4.6, Line 16
- not cleaned
- not random
- entire study
- all words
- weighted

ratings of the lists of words. This is surprising as weighted lists should help focus attention on the most agreed-upon words from the group.

However, this phenomenon is more salient with larger amounts of data, and the relatively small weights (the maximum would be equal to the size of the group - 1) may have contributed to this null result. The similarity scores were 5s, and the range of scores collapsed in the same manner and at the same

Figure 4.13: TurkSim - Study - All Words - Weighted

speed as the unweighted presentation.

4.6 AlgSim

The ratings in this section were generated using an automatic technique that characterized, filtered, and then compared lists of words to one another. Using the method described in Section 3.5.4.2, the ratings here are devoid of human interpretation, save the prior human scoring and interpretations embedded in the WordNet database. They are on a 0-1 scale and cannot be directly compared to the human-generated Likert scale scores of 1-7 from the last two sections.

4.6.1 Raw

The next three graphs are generated by running the study data through AlgSim without any post-study cleaning. This data is raw from the database

and represents what the study participants were seeing displayed in the interface both for themselves and for others. The study software cleaned the incoming data from the HTML text boxes by lowercasing, changing spaces to underscores, and removing other punctuation. No attempt was made to remove typos or plurals or anything else at this time.

4.6.1.1 Random

The ratings in this section are representative of running random data through the AlgSim function. This serves as a baseline for the automatic ratings that follow. The ratings are effectively zero (Figure 4.14).

from Table 4.6, Line 5
- not cleaned
- random
- entire study
- in WordNet
- unweighted

This is expected as the randomness involves selecting random participants and comparing their lists against one another. Any correlation could be attributed to similarity among members of the same work group using the same words to describe their areas of expertise (as compared to participants from other groups, in other industries).

As shown in Table 4.16, the random data for both the "all" words and the "common" words scenarios begin to look statistically significantly different from the study data after Round 1 with an alpha of 0.001. Also listed in Table 4.16 are the same two scenarios, but performed with the hand-cleaned AlgSim
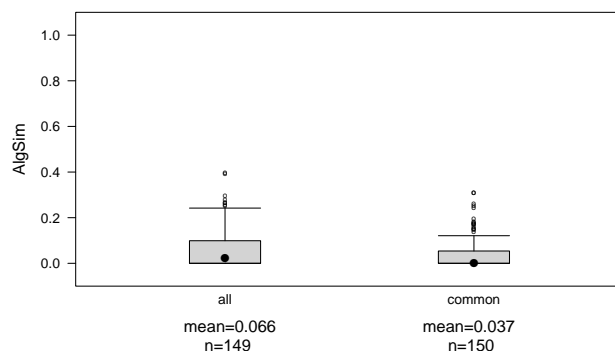
Figure 4.14: AlgSim - Random

raw/cleaned	all/common	Round 1	Round 2	Round 3	Round 4	Round 5
raw	all	0.095	0.000***	0.000***	0.000***	0.000***
raw	common	0.135	0.000***	0.000***	0.000***	0.000***
cleaned	all	0.122	0.000***	0.000***	0.000***	0.000***
cleaned	common	0.046*	0.000***	0.000***	0.000***	0.000***

Table 4.16: AlgSim vs. Random: ANOVA p-values by Round

dataset (see Section 4.6.2.1).

4.6.1.2 Group

The ratings in this section (seen in Figure 4.15) show the group and round level scoring of similarity of the words used in the study. When looking at the common words only, Groups 3, 7, 8, and 9 appear to rise over time. The other groups struggle to rise

from Table 4.6, Line 9
- not cleaned
- not random
- by group
- in WordNet
- unweighted

from zero. When all the words are used, Group 2, 5, and 10 seem to climb a bit as well.
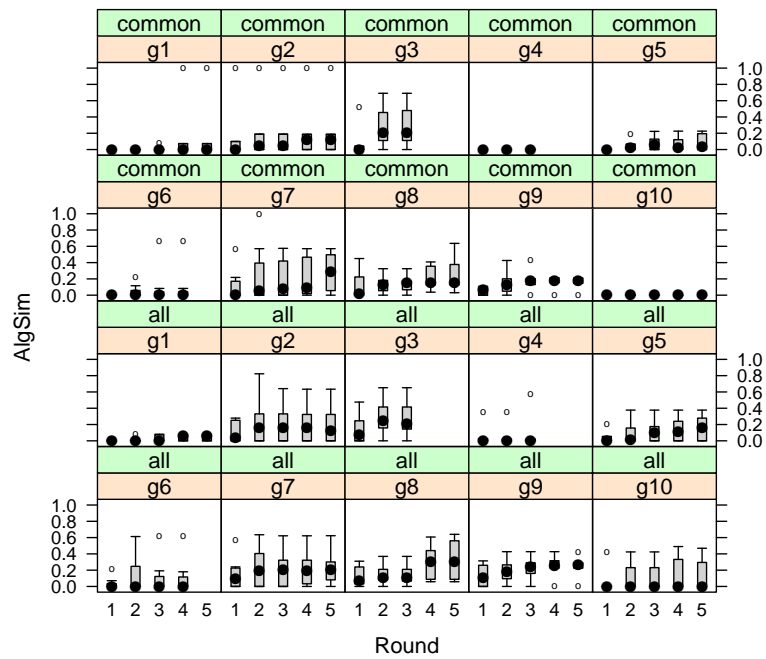
Figure 4.15: AlgSim - Group

Remember, Groups 3, 4, and 6 did not complete all five rounds and therefore have some missing data.

4.6.1.3   Study

The ratings in this section (seen in Figure 4.16) are a composite of the last section and represent the entire study data. The scores rise slightly from Round 1 to Round 5 in both the scenarios where all the words are evaluated and when only the common words are evaluated.

from Table 4.6, Line 13
- not cleaned
- not random
- entire study
- in WordNet
- unweighted
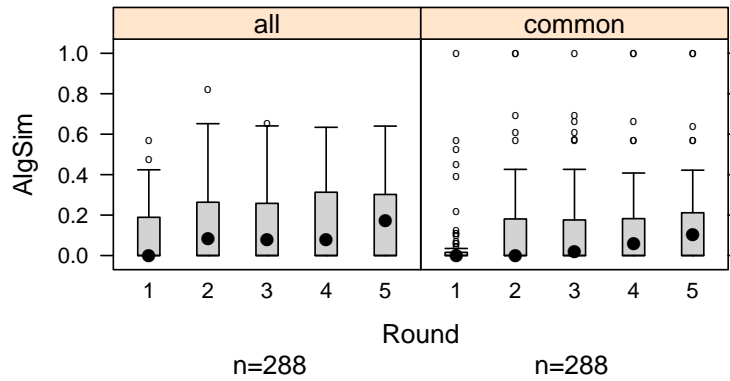
Figure 4.16: AlgSim - Study

A repeated measures analysis of variance (Tables 4.17 and 4.18) shows that the change over time is statistically significant for both scenarios with an alpha of 0.01 (p-values of 0.0036 and 0.0034, respectively).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Round	1	0.30	0.30	8.62	0.0036
Residuals	286	9.92	0.03		

Table 4.17: ANOVA: AlgSim by Round, all words

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Round	1	0.40	0.40	8.74	0.0034
Residuals	286	13.23	0.05		

Table 4.18: ANOVA: AlgSim by Round, common words

4.6.2   Cleaned

The following set of three graphs has been run on the same data as before, but after it was hand-cleaned to remove typos and plurals that did not create confusion and to break multiple ideas into individual words. The hand-cleaning affected roughly 10% of the entered data.

4.6.2.1   Random

The similarity of comparisons of random pair-ings remained very low after the data cleaning (Figure 4.17). The technique used here is the same as before and serves as a baseline against which to measure the next two sections.

from Table 4.6, Line 21
- cleaned
- random
- entire study
- in WordNet
- unweighted

As previously seen in the second half of Table 4.16, the random cleaned data comparisons were significantly different (with an alpha of .001) from the study cleaned data comparisons after Round 1. This shows that the study

data was exhibiting some relevant signal.
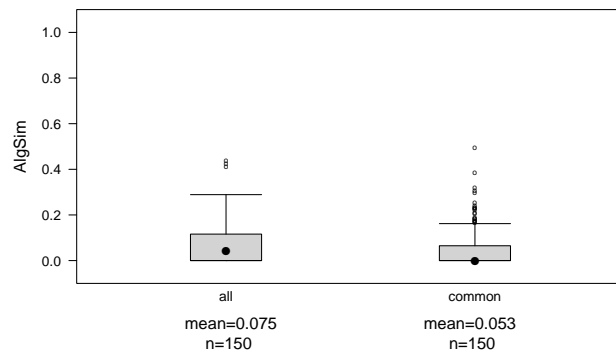
Figure 4.17: AlgSim - Cleaned - Random

4.6.2.2 Group

This section shows the algorithmic similarity ratings by group applied to the study data after it had been hand-cleaned (Figure 4.18). The data has been run through WordNet and words that did not appear were dropped.

from Table 4.6, Line 25
- cleaned
- not random
- by group
- in WordNet
- unweighted

Like the earlier analysis, a few rounds appear to rise over time when the words appear more than once (common). Then when all the words are used, nearly every group shows an increase in their corresponding similarity rating
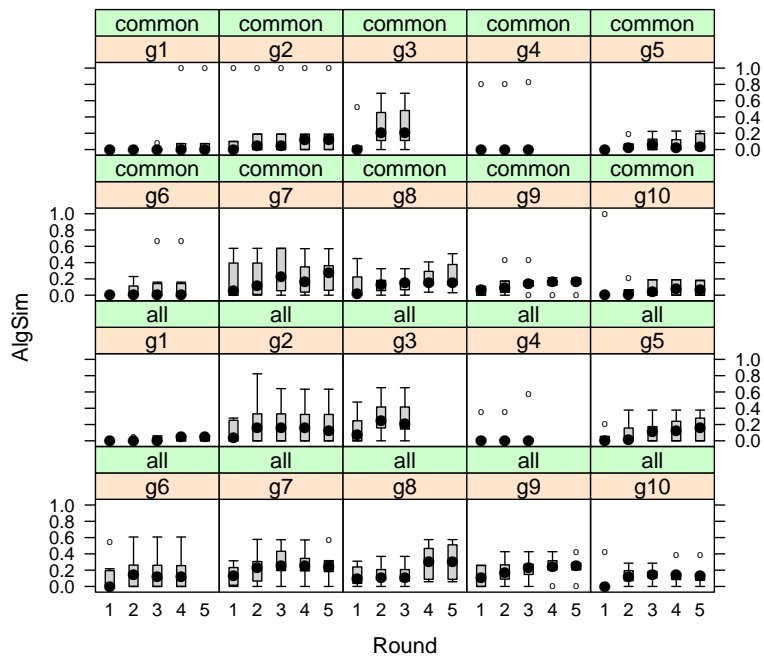
Figure 4.18: AlgSim - Cleaned - Group

scores from Round 1 to Round 5.

4.6.2.3 Study

When the data from all the groups are combined (Figure 4.19), there is a clear rise in similarity score from Round 1 to Round 5. This is the case for both evaluation techniques, when all the words were evaluated and when only words appearing more than once were evaluated.
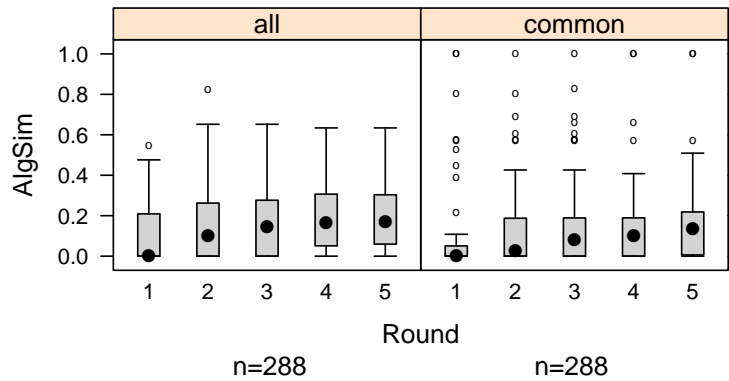
Figure 4.19: AlgSim - Cleaned - Study

The repeated measures analysis of variance (Table 4.19) of the algorithmic similarity scores across time (represented by Round) for the case where all words were evaluated showed a statistically significant main effect with an alpha of 0.01 (p-value of 0.0015).

from Table 4.6, Line 29
- cleaned
- not random
- entire study
- in WordNet
- unweighted

However, looking at a round by round post-hoc ANOVA analysis when all the words are used (Table 4.20), the only significant differences appear to

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Round	1	0.32	0.32	10.22	0.0015
Residuals	286	8.85	0.03		

Table 4.19: ANOVA: Cleaned AlgSim by Round, all words

occur after the original feedback loop between Rounds 1 and 2 (p-values less than 0.05) when the participants are initially faced with their group members' feedback. No other single round or span of rounds show an effect.

	Round 1	Round 2	Round 3	Round 4
Round 2	**0.0249***	-	-	-
Round 3	**0.0058***	0.6428	-	-
Round 4	**0.0016****	0.3906	0.6783	-
Round 5	**0.0014****	0.3480	0.6062	0.9072

Table 4.20: AlgSim - All: Post-hoc ANOVA p-values by Round

When comparing the algorithmic similarity scores across time when only the words appearing more than once were evaluated (Table 4.21), we again see a statistically significant main effect with an alpha of 0.05 (p-value of 0.038).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Round	1	0.22	0.22	4.36	0.0377
Residuals	286	14.50	0.05		

Table 4.21: ANOVA: Cleaned AlgSim by Round, common words

But this time, the post-hoc round-by-round ANOVA shows no significant changes when only the common words are considered (Table 4.22).

	Round 1	Round 2	Round 3	Round 4
Round 2	0.3746	-	-	-
Round 3	0.1152	0.4553	-	-
Round 4	0.1446	0.5035	0.9635	-
Round 5	0.0642	0.2549	0.6522	0.6334

Table 4.22: AlgSim - Common: Post-hoc ANOVA p-values by Round

4.7 Survey

The survey for this study was designed to help evaluate correctness, newness, and potential applicability of the proposed technique for identifying areas of expertise of group members.

The questions in Tables 3.1, 3.2, and 3.3 were issued to all 64 participants at the completion of their group's participation. Eight participants did not complete the survey. Two of these participants began, but did not complete, the survey. The following data represent the 56 completed surveys.

The following ratings (Table 4.23) come from a 7-point Likert scale representing agreement: 1=Extremely Disagree, 4=Neutral, 7=Extremely Agree. The original items are scored directly. The scales are aggregate ratings, representing the average rating for each scale's constituent questions (from Table 3.3).

All but one of the original items scored with mild to strong agreement. The highest ratings of agreement were received by the statements regarding comfort and familiarity of the group members with one another's areas of expertise. Additionally, nearly all participants rated this to be an interesting exercise.

Slightly lower ratings were received by the items regarding the results of the exercise. The participants believed the system gave them somewhat good

and new information that they found useful. They also thought the system did not necessarily gather all the important areas of their expertise and that they would not necessarily use the information to help them make decisions moving forward.

At the bottom of the ratings was the item concerning anonymity. These were fairly small groups, so anonymity is probably too strong a word to use. However, the participants said they would not be as comfortable if the tags had been attributed to one another when being shown. The ability to describe each other's areas of expertise from behind a curtain of deniability increased their level of comfort during the exercise.

Original Items	Average Rating
I am comfortable with my group's tags about my areas of expertise.	5.439
I am happy with my group's tags about my areas of expertise.	5.351
I am familiar with my group members' areas of expertise.	5.333
This was an interesting exercise.	5.196
My group members are familiar with my areas of expertise.	5.175
My group did not list important areas of my expertise.	4.764
I am confident that this system gives me new information.	4.696
This was a useful exercise.	4.679
I am confident that this system gives me good information.	4.643
I am willing to incorporate output from this system into my decision making.	4.607
I would be more comfortable with my group's tags if the tags were not anonymous.	3.298
Scale	Average Rating
Data Quality	4.709
Effort Expectancy	4.670
Result Demonstrability	4.299
Facilitating Conditions	4.250
Performance Expectancy	3.836
Relative Advantage	3.742
Anxiety (reverse coded)	3.036

Table 4.23: Survey Scales and Ratings

Being averages, the aggregate scales are relatively mild and all fit between 3 and 5, straddling the Neutral rating. However, they showed similar results to the original items. At the top of the list, the participants believed this exercise

provided good data quality and was easy to use and clear to understand. The participants rated the items regarding the results of the exercise and its fit within their organization slightly higher than neutral. There was slight disagreement with the items concerning the usefulness and effectiveness of the exercise, as well as whether it would help them do their job better than they can without it. The lowest rated scale item is reverse coded and shows that anxiety about using this tool was not very high. When reversed, the anxiety scale is the most strongly rated set of items (it would score 4.964).

Favorite Part	Response Count
thinking about specific strengths of others	12
what people thought of me	11
more awareness	8
seeing others' self claims	8
how others see others	7
good to reconnect	2
self assessment	2
making connections / learning about others	2
thinking about friends / uplifting / feel better	2
non-job related interests	2
not time consuming	1
similarity and consensus	1
got to know people faster	1
tag clouds of expertise	1
the challenge of listing explicitly	1
help learn about colleagues, otherwise limited contact	1

Table 4.24: Reported Favorite Part of the Exercise

The participants were asked to name their favorite part of the exercise (Table 4.24). The two most favorite things about the exercise were the effortful thinking of others and finding out what others said about oneself. In the next grouping, participants liked having a better understanding and awareness of

their group members as well as seeing how everyone else rated themselves and each other. Mentioned once or twice were more emotional items such as the feeling of reconnection and making others feel good about their skills.

Least Favorite Part	Response Count
redundancy of multiple rounds (3 was enough)	29
nothing disliked	4
yet another email / feeling of tardiness	2
phrasing of tags is hard	2
non-uniformity of terms	2
talking about myself / "not very modest"	2
everyone has a different view	1
trying to determine whether someone was an expert	1
could not go back and modify	1
being asked if i was sure	1
no semantic equivalence	1
stressful	1
vulnerability	1
when others did not reciprocate	1
nervous	1
defining "expertise"	1
fear of future reduced group dynamics because of exclusion	1
realizing i know very little about 3 group members	1
concern over "doing it wrong"	1
entering passcodes manually	1

Table 4.25: Reported Least Favorite Part of the Exercise

When listing their least favorite part of the exercise, the participants spoke with a louder collective voice (Table 4.25). With a resounding 29 responses, the participants really did not like how many times they were asked to complete the same task. They found they were getting little from the group, and had little to add, after the third round. This was expected, fits with earlier research, and probably diluted some of the value of the remaining responses as the participants were only asked to name their single least favorite part. Two

participants responded that they did not like the sense of guilt and tardiness of being reminded they had not participated yet. Two each mentioned that the phrasing of tags is hard and that they were frustrated in how their group did not converge on shared language fast enough for their liking.

Other responses included feeling vulnerable, nervous, having stress at thinking about others, as well as not seeing their self-claimed areas of expertise reciprocated in the tags coming back from the list. Some responses were very specific to the implementation and study-design such as no semantic equivalence being applied to the tags, not being able to go back and modify their tagging within a round, being asked if they were sure before submitting their input, and having to enter passcodes manually.

4.8   Interviews

The semi-structured interviews consisted of the 11 interview questions in Table 3.4. These were issued to 15 volunteer participants from 8 of the 10 groups over the phone and recorded. The average duration of an interview was 24 minutes.

In large part, the interviews reflected similar sentiments as those seen in the survey data. Participants thought the exercise was interesting, easy, fun, provided excellent opportunity for reflection, "reinforced what we already knew" and provided validation, but provided marginal new information and "had too many rounds." They felt nervous and "a bit intimidated" at the beginning but also noted that those feelings dissipated after the second and third rounds.

One thing made more clear in the interviews was the reasoning behind the widespread professionalism of the tags shared within the groups. The exercise was conducted in the context of being a group member, and so they filled the requirement and completed the task as a member of that group, with its constituent assumptions, hierarchies, and efficiencies (or lack thereof).

The interviews made it clear that the open-endedness of the task was moderately confusing and should perhaps have been made more focused. Participants in the role of manager wanted to know how the exercise could be directly applied to task assignment, while those participants who were not managers wanted to know how all the data would be used. A common theme was that they "wanted something more at the end."

There was a clear consensus around anonymity being important within this size group and within organizations where people naturally behave (since they work together). It was thought that removing the anonymity, attributing the tags to specific others, would increase the amount of "sucking up" and lower the signal to noise ratio of insightful and interesting tags. It would also increase the stress and anxiety within the group. Some suggested the total volume of tags would increase, as obligation would become the dominating social motivator. Others suggested it would lower the total volume of tags, as participants would not want to go on the record speaking about one another's areas of expertise. This effect would probably be determined mostly by the individual personalities of the group's members and its existing office culture.

There was the observation that the Self tags tended to be more reflective, conceptual, and considered. The Group tags were more specific, focused on

178

the current tasks that each participant performed at work. A few participants mentioned that talking about oneself was "weird," "awkward," or "advertisey," but also said that "I want people to know more about what I'm doing." One said that he "learned a bit about how I like to be viewed by others" and that he now has "a better understanding of what I want to project." Another said "I need to be better about promoting."

By and large, interviewed participants would recommend this exercise to other groups, but would suggest the focus be made more explicit up front, discussion within the group should be encouraged both during and after, and the results should definitely be shared with new group members when they join to increase their familiarity and rate of acceptance.

The definition of "expert" and "expertise" seemed to be a heavy subject and one that directly influenced participation. Groups that discussed their shared understanding of the concept felt they got more out of the exercise. Those that did not talk as much about the process thought "it was beneficial," but that "it would be more beneficial if we talked about it as an office."

Chapter 5

Discussion

The results presented in Chapter 4 allow the research questions posed in Section 1.7 to be answered directly. In this chapter, the research questions are addressed and analyzed.

5.1 Research Questions and Analysis

5.1.1 RQ1

The first research question addressed by this research was concerned with the performance of Contextual Authority Tagging and whether it successfully identifies a group member's areas of expertise.

R1. Does CAT work?

  (a) **Similarity** - How similar are a group member's opinion of his/her own areas of expertise and the group's opinion of his/her areas of expertise?

(b) **Convergence** - How does the similarity behave over time? Do the two opinions converge? If so, how long does it take? If not, is there a persistent gap?

Since expertise was operationalized as a set of attributed words, or tags, similarity of expertise was operationalized as a rating comparing two sets of those words. A high similarity rating suggests the two sets of words were covering the same ground and had little impedance mismatch.

Both of the human-based similarity techniques suggest that the proposed process succeeds in providing group tags that are similar to an individual's self tags. The automatic technique gave scores that were relatively low on its absolute scale, but statistically significantly different from random comparisons.

The results from the automatic WordNet-based approach suggest that the algorithm could be made to better agree with the human assessments of similarity. The algorithm itself was relatively simple and did not employ any sophisticated data-mining, word stemming, or semantic equivalence judgments in its calculations. With a bit of training and semantic capability, the automatic approach should improve dramatically and bring the effort and cost of this type of analysis down significantly from the two human-based approaches.

In order to discuss convergence of a subjective evaluation, I chose to employ an increase in similarity as a proxy for convergence. Both the TurkSim and AlgSim techniques did display a rise in similarity over time and can therefore be said to approach convergence. It would be too bold to suggest they ever converged, as the richness of language can never be fully characterized by compressing it into some subset of constituent words. Additionally, the groups

in the study were particularly scoped and did not venture much beyond the job tasks the group members are assigned and carry out at their workplace.

WordNet was definitely an enabling technology as well as a limiting agent during this process. Without the extensive prior work done to categorize and quantify the English language, the automatic technique I proposed and executed would not have been possible. However, due to the free-text phrases that my participants regularly used to describe their areas of expertise, many rich contexts were lost due to the necessary dropping of the words and phrases not found in the WordNet dictionary. As a general purpose tool, Contextual Authority Tagging will probably need to have some human curation and/or some discipline-specific dictionaries available to better contextualize and understand the knowledge being shared when group members are describing one anothers' areas of expertise. Expertise is related to depth of experience and understanding and the deeper one goes into a subject, the more specific and context-laden the vocabulary becomes.

The survey data shows that the participants themselves also suggest the exercise provided them with good representations of their own areas of expertise and that of their colleagues. Besides the reverse-coded Anxiety scale, Data Quality was the highest-rated scale. Participants also felt that their group members did a good job labeling their areas of expertise. Slightly less, they agreed with the notion that their group members left out, or did not list, some important areas of their expertise. This suggests that with more data and perhaps more participants, the data quality could improve as coverage (familiarity) fills out and spreads across additional domains.

The first hypothesis from Chapter 1 was:

H1. As the social fact of what a person knows is molded by the group, a consensus will appear and converge.

The data supports the view that Contextual Authority Tagging provides a baseline for concluding that a group's opinion about a person's areas of expertise can give good information. A consensus appeared, was agreed to by the individual being tagged, and somewhat converged over time as the language and norms of the group were negotiated in a shared space. This finding comes with the caveat that the participants knew one another well enough or had enough experience with one another to feel the data being provided was of good enough quality. When conducted outside of well-known groups, this finding may not hold as both participant identity and the promise of future interactions are not as strong.

5.1.2 RQ2

The second research question addressed by this research was about the more practical issue of the acceptability of Contextual Authority Tagging as a new technology.

R2. How acceptable is CAT?

(a) **Comfort** - How comfortable are group members in participating? What are the main factors influencing their comfort level?

(b) **Confidence** - How confident are group members in a system like this? What is the quality of the output of this system? Does this system provide a valid credential? Does this system increase users' trust in one another?

(c) **Usefulness** - What is useful about a system like this? What did participants learn? How would using this system affect participants' decision making?

This turns out to be a hard question to quantify, but generally speaking, the demonstration tested here is not ready for a production environment and should be improved in a number of ways before being used to provide actionable results to any organization.

That said, participants reported that, while they had some early apprehension about having co-workers talk about them, they quickly realized that their co-workers were well-behaved and that a system like this could work because everyone is equally visible and "vulnerable." Sufficient external incentives were already in place, apart from the study software, to make sure participants were cordial and respectful. The participants said they were comfortable with the tags their group members listed about them and that their cautiousness went away after the first couple rounds, as they began to understand the dynamics of this type of a system and its affordances both for themselves and equally so for their colleagues. The visibility and comprehensive simplicity of the bimodal self/group construct played a role in the participants feeling comfortable. They quickly made peace with what others could see. The complexity of the system is in its social application and feedback loops, not in its structure.

The participants in this study also voiced through their survey responses and via the follow-up interviews that their confidence in this system to provide good information was fairly high. They felt confident that what they were seeing was correct, but not comprehensive, and they would assume that because these characteristics were true for their own tags, they would be true for the other group members as well. Multiple participants shared that there was well-known information that was not being shared within the tags, keeping it invisible to the system. They speculated this to be because the well-known information was off-topic for a work-related exercise, and that the person whose areas of expertise were being omitted would not have liked off-topic information being discussed in a work-related tool. This sentiment was most commonly expressed about those in positions of authority, i.e., the boss. While this may fairly represent the existing culture of any particular group, it may actually be representative of a failure to capture some topic or area that could prove important later. Willingness to describe and be described outside of the professional job description would be an interesting independent variable for future research into productivity, teamwork, or morale to consider.

Regarding the question of being a valid credential, the participants agree that CAT could serve as a low-to-mid value credential that would be trusted. It would point them in the right direction, but they would want to verify any information that was suggested by the group that they had not personally experienced before relying on it. This new unverified, but community-sourced, information could be seen in the same way that Ronald Reagan described the United States' working with the Soviet Union when emerging carefully from

the Cold War, "Trust, but Verify." A few participants referred to "learning" something about someone, striking up a conversation later and being pleasantly surprised that the item was true. One female near-retiree was surprisingly tagged with "swordfighting." This was met with incredulity by her other colleagues until more details were provided face-to-face to her group members. It simply had not come up at the office prior and the explicit nature of this tool provided an opportunity for discovery and discussion.

While nearly all the participants thought the exercise was interesting and sometimes thought-provoking, their sense of its usefulness was moderated by the fact that they were not sure of its goal or what they should be getting out of it. A few mentioned small things they learned about a colleague, but were not sure how it may help them in the future. All of the "new" information discussed in the interviews were not job related and were considered "trivia" about the colleague, not directly influential on their duties or capabilities. However, where the roles in an organization or group are not as formalized and more fluid, or when the organization itself is less role-based (e.g., a non-profit or volunteer organization, or a consultancy), participants thought that this type of information could be useful to help uncover unknown talents or histories. Additionally, multiple interviewees stated that this type of information about their colleagues would have been very useful to see when they first started their job; they would have liked to have had a quick tagcloud-like overview of each person in the group when first orienting themselves to a new workplace. They expressed that having an asynchronous tool to learn about others would be very helpful. Knowing that anyone within the community could edit it, and

that others were paying attention, would warrant it additional trust.

However, none of the participants mentioned that this type of a tool would influence their decision-making within their role as a group member in the context of their workplace. I suspect this is because this keyword representation is a "watered down" version of what they know more richly in person. I suspect also that CAT's ability to influence decision-making would increase somewhat if the relationships being evaluated were not as rich; without existing personal interaction and history from which to sample, the output from CAT would become a better relative source of information. An interesting corollary suspicion is that with a more personal knowledge and history shared among group members may come a relatively less compelling set of tags. Groups that know each other well may not learn as much from a tool like this simply because everything is well-known already. But then again, the value may be greatest for those outside the circle, those who are not already well-known to the group.

The second hypothesis from Chapter 1 was:

H2. Comfort levels will increase as the system becomes known and understood. Initial trepidation will be assuaged as the system allows participants to see more of how they are perceived by others.

This hypothesis was supported both by the survey data reporting low anxiety as well as direct statements from interviewed participants. They realized how the system was sharing their inputs and they experienced seeing the inputs of their colleagues concerning their own areas of expertise. Participants stated very clearly that they became more comfortable over time.

The third and final hypothesis from Chapter 1 was:

H3. Group members will have confidence in this system and exhibit increased trust in one another.

This last hypothesis was found to be partially supported. Participants did have confidence in the system to collect and then report the type of information they were expecting it to report. They thought the data would be quality data and they trusted it for what it was.

However, they did not report that the trust in the data carried over to increased trust in the other participants. The study design forced the group members to already be acquainted with one another and have existing working relationships. This means that the participants began the study with a fairly high degree of trust. This study provided no support for the idea that participants' trust levels increased because of the exercise.

It would be interesting to ask a specific set of questions about colleague trust of a set of group members who were just beginning to work together or of group members who knew each other in a less formal environment than their salaried jobs.

5.2 Conclusion

After considering the group results, the survey and interview responses, and the analysis above, I think there are three major points to take from this research.

The first conclusion is that Contextual Authority Tagging succeeds in identifying the areas of expertise of group members, within a workplace of trusted

peers, but should not be considered a standalone technique or process. This type of software framework should be integrated with existing email, personnel, or people-centric technology and provide a meta-layer above and beyond what is already being provided. A separate product is too cumbersome and disjoint from an employee's already-full business workday and is less useful than having a simple context-aware overlay available to them in a familiar tool or environment. As long as the input interface remains just a text box and the output is just a ranked list or tagcloud, the mental model of self/group is simple enough to include within other interfaces without fear of introducing any overwhelming complexity or clutter. Having a larger company's human resources department administer this type of a system would probably make the most sense as their primary unit of focus is already the individual worker.

The second conclusion is that if Contextual Authority Tagging hopes to provide a service or improve communication within a business environment, it needs to do a much better job of providing some guidelines for interpretation in addition to simply providing data. I had assumed that providing a new set of information about an interesting topic (data about one's self and colleagues) would be enough to encourage discussion and reflection. Unfortunately, hardly any participants reported talking about the collected and reported data within their groups despite nearly all reporting that it was very interesting to see what their colleagues thought about them and each other.

A tool like CAT should be deployed in a workplace environment with clear explanations for each of the questions: who, what, when, where, and why. Without these types of background agreements in place, new technology has a

much harder time getting buy-in from its target audience and since the target audience is being paid to do their jobs, it needs to be framed with respect to those jobs. Groups will get the most out of a system like CAT if they take time to collectively and consciously reflect on their definition(s) of "expertise", what they hope to get out of the exercise, and the terms being used to describe the different members' areas of expertise. Additionally, if this exercise is part of a consulting service offered to a group, some interpretation should accompany the results data. Some of the most powerful reactions I heard in the interviews were from individuals who had gone through these framing and interpretation steps of their own volition. Groups that did not discuss the exercise amongst themselves were distinctly less engaged with their results.

The third conclusion is that CAT does not generally provide "new" information to small, well-knit groups of co-workers. That said, group members who were newer or less connected to the "core" of a group responded that they did learn a bit about their colleagues. They speculated that they would have learned about their peers much faster had this type of a system been in place when they arrived. Well-known information may not be as exciting to established members of a group, but that same known information is extremely useful and interesting to people who are not established members of the group. And since this type of information is "vetted" by the group, it holds a high degree of validity to outsiders and new entrants.

Chapter 6

Conclusion

In this final chapter, I share a summary of findings and confirmations of earlier work, explain the contributions and implications of this work, discuss a few lessons learned and the limitations of the particulars of this study, and share some thoughts on practical future research directions based on Contextual Authority Tagging.

6.1   Summary of Findings

Overall, this research has provided insight into how familiar groups of individuals in the workplace can understand what their colleagues think of their areas of expertise. This work has shown that, with simple keywords, group members can convey the salient areas of expertise of their colleagues to a degree that is deemed "similar" and of "high quality" by both third parties and those being evaluated.

Identity formation and negotiation is alive and well, and this research fits within the frames drawn by Goffman (1959) and Tajfel and Turner (1986) and

furthered by boyd (2002, 2008). We perform and we understand ourselves in part by understanding the reflections that come back to us from others (Marchionini, 2009).

In a fast-moving networked workplace, this ability to gain insight into the knowledge of others with a simple trustable lookup may prove valuable. Tapping into the collective understanding and distilled opinion of those around us could be a useful tool or sanity check against both direct and indirect individual claims of expertise. Equally, it could serve as a weapon against misplaced modesty, allowing us to collectively reward those who deserve to be given credit when credit is due.

What remains an open question is whether this type of collective opinion mapping works in an environment beyond the walls of the relatively small, trusted workplace, where people know one another (stable identity) and have many incentives to behave and only say positive, professional things about one another ("the shadow of the future" (Axelrod, 1984)).

6.2 Contributions and Implications

One of the surprising results and potentially important contributions is that weighting did not seem to affect the similarity ratings of two lists by third parties. Both the HumanSim and the TurkSim ratings were largely unaffected by the addition of weights (occurrences) to the words in the compared lists. This was unexpected and probably deserves some further investigation. The public's continuing affinity for relatively-sized tagclouds suggests there may

be something missing in the current study design that suppressed this affinity. The size of the lists were not terribly long (it was possible to keep the entirety of a long list within working memory) and this may have contributed to the weightings' apparent lack of an effect. With longer lists and more data, the weightings may have proved a more useful proxy or shortcut and therefore added more value.

Another interesting observation is that, within this study, expertise sometimes appeared to be understood as existing on a spectrum ranging between sets of skills and individual interests. Participants appeared to use this spectrum of understanding by discussing the group results in terms of the two extremes, one facing "forward" and one facing "back." One manager spoke about the two lists with an added layer of (self-imposed) meaning; listing things for one's self was showing interest or desire to work on certain tasks ("forward"), while the group lists uncovered what others thought that person was good at and what their current job descriptions were about ("back"). This dichotomy allowed the manager to infer where new projects may better fit with the existing personnel's interests in addition to fitting with existing forged expectations and demonstrated prior experiences.

I was surprised by this layering of intention and meaning on an otherwise straightforward request to list areas of expertise. I remain unconvinced that this meaning is well-founded and that it may speak more to a lack of equivalent communication (concerning intent/desire) happening elsewhere within these groups. It feels backwards to divine intent or desire when intent is one of the easiest things to ask directly of an employee. Again, the self assessment part of

this tool was largely provided as a check on the group feedback. The fact that this layering of meaning appeared in multiple conversations was interesting and warrants further study from a managerial perspective.

I think one of the most fascinating findings from this study is that one of the reported "most favorite" parts of participating in this study was the effortful thinking about the specific strengths of others. Apparently, people really liked the process of thinking about the positive aspects of those around them, those with whom they work. They reported that this process was invigorating and made them feel good. I would assume that this effect would dissipate if they did not perceive their feedback was being viewed or read by the subject. I expect there is significant opportunity in this area for morale and performance research to contribute further insight.

On a more theoretical level, some confirmation of the effects modeled by the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003) was also demonstrated. The model, which incorporates eight earlier models including Social Cognitive Theory (SCT) (Bandura, 1986), the Technology Acceptance Model (Davis, 1986), and an instrument for measuring Adoption of Information Technology Innovation (Moore & Benbasat, 1991) predicted that anxiety would not be a significant influencer over time. The results of this study bore that out and suggest, with experience and familiarity, anxiety does not have a significant effect on intention of technology use. Additionally, there was no suggestion from the participants that social influence had a large role to play on their participation or their sense of obligation to the study. Venkatesh et al. (2003) explicitly stated that when a technology is

voluntary in an organization that social influence would not be a significant factor. This validates my excluding this factor from the survey as this study was completely voluntary in nature and would presumably be deployed as such in any workplace environment.

Additionally, Venkatesh et al. (2003) showed that the UTAUT model explained 70% of the variance of usage intention. The strongest predictor within the model is Performance Expectancy and this study had the Performance Expectancy of CAT scoring the lowest. It also was the only scale item to score below Neutral, suggesting that CAT does not meet a standing demonstrated need or provide an apparent benefit to the participants at this time.

Coupled with the strong findings regarding Data Quality and Comfort (lack of anxiety) and strong feedback in the interviews, I feel there is a temporal aspect to the UTAUT model that might currently be missing. The interviews suggest a more direct relationship between having confidence in a system and low anxiety before being able to effectively evaluate things like Performance Expectancy, Effort Expectancy, and any Facilitating Conditions. I would suggest a more temporal model where comfort and confidence are both necessary before usefulness can effectively be considered. An alternate interpretation may suggest that UTAUT already considers comfort and confidence a part of Performance Expectancy, in that without them, a participant would not expect a system to have an effect on their relative performance in an organization.

Apart from the specifics of this study, the reputation framework generated during my comprehensive examination, based on Sabater and Sierra (2005),

has also proven useful when thinking about CAT and its potential future incarnations.

6.2.1 Reputation Framework

The following seven components cover the most interesting and salient features of most reputation systems.

1. **Unit** : First, any system must be measuring the reputation of a particular unit. What this unit is determines much about how the system will be realized. Some systems are based on people, others on documents, and others on organizations or the reputation claims themselves.

2. **Global/Local** : That said, most important, from a design perspective, is whether the system is calculating a global or localized reputation "score." Whatever is being calculated, if it is to be a representation of an object from the perspective of "the system," then it must be globally shared and accessible by the entire system. This is most easily done in a centralized system where a single codebase or algorithm is both determining a score and storing it for further access and distribution. A distributed system is much more complicated to engineer and police, but also, more robust in the case of failure or infiltration.

3. **Algorithmic/Cognitive** : A third criterion would be the nature of the information being stored – whether it is of a deterministic, algorithmic nature or cognitively generated. This component is important as to the level of interaction humans must have with the system to run the

system. As a completely automated, generative algorithm can be run many times, very quickly, it may be missing the ability to change over time in the ways that a human-based system may find natural or easy.

4. **Recursivity/Transitivity** : A fourth component of reputation systems is their level of recursivity and transitivity. Systems that are global may integrate some amount of dampening or multiplication within an algorithm when determining reputation scores, but have little with regards to recursive programming. Distributed systems usually require participating actors to communicate amongst themselves both their scores and some additional information regarding "hops" or "TTL" (time to live).

5. **Direct/Summary** : Some systems take measurements on one type of item and use those to calculate a score for some other item, as an aggregation or summarization score. Other systems are actually evaluating the item of record directly. Direct systems are much easier to conceptualize and follow algorithmically, but may provide very simple outputs and less insight into the nature of the thing being evaluated. More complex models are harder to get right, but may convey more meaning when they produce useful output.

6. **Transparency** : Transparency is a component that is key to understanding a reputation system. The double-edged sword, of course, is that, with transparency, insights into the ways that calculations are made comes clarity into how to game the system and provide unfair advantages to some over others. Transparency is usually evaluated on a spectrum,

as having full or no transparency is rarely the best option.

7. **Reliability/Confidence** : Lastly, some amount of internal scoring and ranking may happen between nodes of a system to help calculate confidence or reliability. If a system can be infiltrated, the individual participants must have a way to ignore or punish misbehaving actors or the system will become completely useless very quickly.

As an illustrative exercise, the mapping in Table 6.1 shows the above framework applied to Slashdot's Karma system for managing online comments, Google's PageRank algorithm for evaluating the relevancy of web pages to search terms, and BitTorrent's protocol for handling the behavior of nodes that comprise the peer-to-peer filesharing network.

Component	Slashdot's Karma	Google's PageRank	BitTorrent
Unit	account (person)	document	peer node
Global/Local	a single score for each account	a single pagerank score for each document	scores are calculated at the node level and shared openly
Rational/Cognitive	cognitive (human)	strictly algorithmic	strictly algorithmic but with knobs available to node operator (human)
Recursivity/Transitivity	scores are not propagated between accounts	high value documents propagate value to their linked documents	scores are reported during discovery
Direct/Summary	based on comments	based on links	direct observation
Transparency	metamoderators can see moderation data and the underlying code is open	specific code is closed but the basic algorithm is well known	code is open and flowing data is completely visible
Reliability/Confidence	confidence based on a few trusted humans	based largely on result quality	high confidence based on visibility into the data

Table 6.1: Reputation Framework: A mapping of components and systems

6.3   Lessons Learned and Limitations

The first thing I would like to share here is more of a lesson remembered. I have always felt that I am a toolbuilder and I am always pleased to share tips and tools that I use with others (see Appendix 6.5). Very early in the process of writing this dissertation, I realized I would need a system for writing this document that did not require me to go back later and reformat or recalculate anything manually. I probably spent too much time automating and scripting the tools that I used to generate the tasks, assign the tasks, retrieve the results, organize and manage the results, and then analyze and display the results. But with the time spent, I gained confidence in the process and feel that I came to learn about the data in a more intimate way.

The second lesson learned is related to the difficulty I had in recruiting participants for this study. My committee warned me that finding organizations to work for me would be more difficult than I had suggested, and they were correct. I vastly underestimated the time and effort it took to recruit and manage ten different groups. Groups declined to participate due to time, due to a "lack of fit" with their organizational goals, and due to current office politics and/or morale. In any future work related to this type of group-based evaluation, I would definitely suggest having a short demonstration available; a hand-holding exercise for the first round. An example from a previous group would also suffice (with granted permission, of course). A full case study of how this type of evaluation helped another similar business would probably be most helpful for future recruitment efforts.

One of the first limitations of this study is also related to the difficulty

in recruiting groups of participants. Groups that did respond generate some level of a self-selection bias which I would guess leans towards friendly, positive, and well-behaved. There is no good way to ascertain the reasons behind non-response from potential participants as, naturally, they were not responding to the initial snowball request for participants. It would be interesting to do a more in-depth follow-up with a few of the initial individuals that did not respond at all.

A definite limitation of this current work is the small size of the groups involved. I was pleased with the overall group sizes (averaging 6.4 participants per group) but was originally envisioning groups of 8-10 (which would have provided roughly twice as much data about each participant, i.e., $9^2 = 81$ assessments rather than $6^2 = 36$ assessments). Looking back, it makes sense that most corporate group sizes are not quite that big. Most feedback suggested that once a group was 8 or larger, it was broken into two groups of 4 or 5 as that makes for less overhead and quicker, more agile, decisions and meetings. As such, this small group size probably had an effect on the sentiments expressed about the lack of "new" information coming from the tags. It suggests that, if a group was going to see any significant rates of "learning new information," the group would need to be at least large enough that the participants did not feel so familiar with their group members. Group members would not necessarily know enough about everyone to tag everyone, and this would be by design. Using relatively small groups also limited the study's ability to predict how this type of exercise would function in a larger context (at a unit or company level). I suspect that more "new information" would begin to appear

one organizational level "up," where a group of groups participate together to better learn about those they work "near" but not "with." I hope to explore this aspect of CAT in the future.

Another possible limitation is related to the backgrounds of the individual similarity raters. In both the HumanSim and the TurkSim similarity comparisons, the raters were not subject-level experts themselves. This may have affected their ability to evaluate any possible in-group language usage from tightly coupled groups with specific or technical areas of expertise. Certainly the WordNet database did not have good insight into engineering phrases or particular areas of legal practice. This would suggest that, when non-subject-area-experts are evaluating this type of data, their evaluations could be biased towards "more similar" than if subject-area-experts were looking at the lists and seeing more nuance. I suspect that, with research of this type with larger groups and more diverse participants, the evaluation problem could shift from one of "too narrow" to one of "too many different levels of detail" collapsing into one namespace.

The last limitation, or potential side-effect, concerns the psychology of participants as they continue working together after participating in this study. If a set of words becomes associated with a particular person, this could have some potential for limiting the views of others who come into contact with the tags out of context. This person may be unfairly pigeonholed in the future, or left out of certain conversations, since their tags did not reflect their interest or expertise in a particular area. This is part of a larger pattern of dealing with a more algorithmically determined environment more generally (Pariser,

2011). As we continue to depend on the products of algorithms as our inputs for daily decision-making, we must keep in mind the bigger picture and our original intentions. If we lose sight of what the data we have represents and the limitations of how it was gathered and processed, we may begin to measure and value the wrong things.

6.4   Future Work

In addition to some of the work mentioned above in the limitations section, this study begs for more detail to be provided across different contexts and different algorithms.

First, the limits of running this study on only 64 participants are paramount. Second, these participants were in a limited physical region of the United States and expanding the scope of this work would provide significantly stronger evidence for a group's ability to assess and provide credentials for its members. In addition, it could provide further opportunities to assess group dynamics in other parts of the country, in specific fields or industries, or with groups of differing sizes. Other variables ripe for study include whether the tags are attributed (anonymity, pseudonymity, etc.), whether the participants are allowed to pre-approve the publication of their group-awarded tags (potential whitewashing/grooming), and different types of groups. Additional types of groups, besides corporate, could be family groups, hobby groups, neighborhoods, social groups, professional affiliations, and just friends. Each of these

would provide rich fodder for further sociological, organizational communication, identity formation, reputation, and interpersonal research.

I would also like to see this type of timestamped tagging analysis done over a longer period of time. The strongest negative response was to the repetitive nature of the study design and the feeling that the participants had to do the same thing over and over again with little return for their efforts after the second and third rounds. This fits with the Delphi Study literature and was not a surprise. However, I think this represents a simple first "discovery" phase which would be followed by other, longer, slower to develop phases if the analysis was continued. This would require a more continuous system without the forced iterations that made some participants feel tardy and guilty for not responding as fast as their group members. If a continuously updated and current system was available through a company's personnel database or directory service, regular additions to the tags could automatically percolate to other systems or reports. If this type of assessment were part of regular performance or job evaluations it could prove cheaper and more efficient than capturing the same information in other ways. A longer-term analysis of expertise data may also afford reporting where company-wide aggregate arcs of knowledge could be mapped and trends could be discovered. As mentioned before, once something can be counted, it can also be graphed over time, modeled, and perhaps even predicted.

Additionally, my human similarity raters brought to my attention their troubles considering differing lengths of lists. I had not considered how two lists, when one is a complete superset of the other, might be interpreted. It

turns out this is a very subjective question to answer ("How similar are these?") and may depend on a variety of interesting factors.

More specifically, what does it mean when the group shares information that is continually not reflected back by the individual? Are they ignoring the signals from the group? Are they trying to downplay a certain skill set? Is it old? Is it passé? Are they embarrassed?

And what of an individual that is continually suggesting they know more than is being reflected back? Is that persistent gap a demonstration of untapped expertise? Or is it a call for attention? Perhaps it is boastful arrogance that is correctly being ignored by the rest of the group? Learning more about what is happening when there is a Self/Group imbalance seems a rich area for continued research.

There is also a large potential for reducing the cold-start issue when creating tagging databases. There is probably a good way to import data (tags) from other data sources to jump-start good discussion and participation. A good data source (or word bank) would be scoped in the desired direction (professional, off-work, medical, etc.) and could have huge effects on how quickly and effectively an organization took to using a tool like CAT. A few exemplars of "model" behavior can go a long way in establishing the culture around any new practice or tool.

As the corollary to better data coming out, a better way to get tags into the system may significantly affect uptake or enthusiasm about a tool like CAT. If tags could be entered via voice capture, copy and paste, or even as a result of some game dynamics, this could lead in interesting directions.

One of the most significant areas of improvement could be found in the automatic algorithm used to determine similarity scores themselves. The Mihalcea et al. (2006) algorithm used here was limited to a bag-of-words analysis and did not have any outside knowledge or context for the words being thrown at it to process. A topical dictionary could be used to retain more words (and even decipher acronyms) for participation in the algorithm. Stemming could be used to compress plurals and other words with similar roots. A clustering algorithm could be used to generate an additional parameter for the similarity calculation.

A much more grand vision for this work is to deploy at web-scale a backchannel for expertise tags to be attributable to web-wide identifiers. This datastore would create a marketplace for different inputs (from web pages, résumés, journal articles, source code, credentials, etc.). It would also create a marketplace for filters and provide some ammunition in the fight against what Clay Shirky has called "Filter Failure" (Shirky, 2008b). Additionally, if authority tagging is successful at a large scale, filtering the quality tags (and quality taggers) from the noise quickly presents itself as a new problem to be addressed. This would in turn create a marketplace for knowledge about the filterers.

As we each need to make explicit value judgments about what information sources we believe have credibility, we would begin to choose certain providers over others. We would make these decisions based on theory, convenience, experience, and recommendation. It quickly becomes a recursive reputation problem – one we have seen before. How does one know whom to trust to help one determine whom to trust? The selection of a filter would mean as much

to this type of an ecosystem as the selection of a particular expert. Having a transparent infrastructure that was able to support this level of determination and record-keeping for those who wish to share and benefit from the collective activity could provide a robust capability to filter the noise from a vast sea of opinions.

If this vision is to come to pass, much work will need to be done in bridging the small corporate groups studied here to a global network of loosely-connected individuals trying to perform the same task of tagging each others' areas of expertise. However, my hunch is that by combining what we have learned about Networks (Barabási, 2002), Small World Theory (Milgram, 1967; Watts & Strogatz, 1998), and The Strength of Weak Ties (Granovetter, 1973), it can be shown that the same effects are possible at web-scale.

6.5   Conclusion

This research into Contextual Authority Tagging needed to be centered in theory and found its home in the corporate knowledge management literature as well as network science, reputation and identity, and classification theory. The intersection of these areas suggest a tightly coupled meritocracy where transparency and trust are possible and expertise can be rewarded.

We now live in an ever-shrinking world of always-on connectivity and powerful communication devices. Since these devices are two-way, they provide a voice (and a distribution platform) to millions who, prior, have never had a voice. This is a remarkable achievement and serves as a testament to the

incredible advance of technology and our collective striving for equality with regards to opinions and freedom of speech. However, with monumental increases in the number of voices and opinions being shared, we demand a requisite increase in the power of tools to help us filter all this newfound information. We need good knobs to help us determine where to direct our always-limited and increasingly precious amount of attention.

The freedom to listen to anyone has to be balanced with the practicality of not being able to listen to everyone. We need tools that help us serve both of these needs, albeit not at the same time. The tools need to be flexible enough to let us listen to whomever, whenever and wherever we want, and to reserve the right to change our minds at a later time.

Finding good sources of information is hard. Knowing whom to listen to when the subject matter is beyond one's personal experience is a daunting and important problem, but one that can be reduced to an engineering problem with the right approach.

By showing that a group's opinion can be quantified, validated, and trusted, I feel Contextual Authority Tagging has taken the first small, but foundational step towards a future with a functional ecosystem of marketplaces for expertise inputs, filters, and brokers. This, in turn, may help us make sense of a democratic world where everyone has a voice.

Appendix A: Colophon

Over the years I spent on this work, I developed some strong opinions about the software and platforms I was using along the way. The unix philosophy of having a set of singular tools that each do one thing well guided my hand and I found the following tools most effective.

The Contextual Authority Tagging software was written in PHP and MySQL. The similarity tools used to process the data were written in PHP, Perl, and bash. Images and statistics were produced with R and OpenOffice. Survey data was captured using Qualtrics, licensed through UNC-Chapel Hill's Odum Institute.

This LaTeX document was written and managed in TextMate on Mac OS X. The references were saved in BibTeX format and managed with BibDesk. The statistics are included with Sweave and generated from the accompanying raw data files. The resulting PDFs were viewed in Skim.

The entire process (text, statistics, scripts, data, notes) was saved along the way in Subversion.

Clearly, I owe a debt of gratitude to the vast array of open source software that made this project possible.

References

Abecker, A., Bernardi, A., Hinkelmann, K., Kühn, O., & Sintek, M. (1997). Towards a well-founded technology for organizational memories. *American Association for Artificial Intelligence Technical Report*, *SS-97-01*. 11, 14

Abrams, J., O'Connor, J., & Giles, H. (2002). Identity and intergroup communication. *Handbook of International and Intercultural Communication*, 225–240. 30

Ackerman, M. S., & Malone, T. W. (1990). Answer garden: A tool for growing organizational memory. *Conference on Supporting Group Work*, 31–39. Available from `http://portal.acm.org/citation.cfm?id=91474.91485` 14

Anderson, C. (2006). *The long tail: Why the future of business is selling less of more.* Hyperion. 91

Apple. (2007). *Security - get a mac ad.* Available from `http://movies.apple.com/movies/us/apple/getamac/apple-getamac-security_480x376.mov` 99

Axelrod, R. (1984). *The evolution of cooperation.* New York: Basic Books. 60, 192

Baeza-Yates, R. A., Jones, T., & Rawlins, G. J. (2000, September). New approaches to information management: attribute-centric data systems. In *Seventh international symposium on string processing and information retrieval (spire 2000)* (pp. 17–27). A Caruna, Spain. Available from `http://dx.doi.org/10.1109/SPIRE.2000.878176` 87

Baier, A. (1986, January). Trust and antitrust. *Ethics*, *96*(2), 231–260. Available from `http://www.jstor.org/stable/2381376` 60

Baker, W., & Dutton, J. E. (2007). Enabling positive social capital in organizations. In J. E. Dutton & B. R. Ragins (Eds.), *Exploring positive relationships at work: Building a theoretical and research foundation* (p. 325). Routledge. 54, 55

Balog, K., Azzopardi, L., & Rijke, M. de. (2006, August). Formal models for expert finding in enterprise corpora. In *Sigir '06: Proceedings of the 29th annual international acm sigir conference on research and development in information retrieval* (pp. 43–50). Seattle, Washington. 15

Balog, K., Azzopardi, L., & Rijke, M. de. (2009, January). A language modeling framework for expert finding. *Information Processing & Management*, *45*(1), 1–19. 13, 15

Balog, K., & Rijke, M. de. (2007). Determining expert profiles (with an application to expert finding). In *Proceedings of the 20th international joint conference on artifical intelligence* (pp. 2657–2662). 15

Balog, K., & Rijke, M. de. (2008). Non-local evidence for expert finding. In *Cikm '08: Proceeding of the 17th acm conference on information and knowledge management* (pp. 489–498). New York, NY, USA: ACM. 15

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ, US: Prentice-Hall, Inc. 194

Barabási, A.-L. (2002). *Linked: The new science of networks.* Basic Books. 38, 206

Barnard, C. I. (1968). *The functions of the executive.* Harvard University Press. 33

Becerra-Fernandez, I. (2000, October). The role of artificial intelligence technologies in the implementation of people-finder knowledge management systems. *Knowledge-Based Systems*, *13*(5), 315–320. Available from `http://dx.doi.org/10.1016/S0950-7051(00)00091-5` 16

Becks, A., Reichling, T., & Wulf, V. (2004). Expertise finding: Approaches to foster social capital. In M. Huysman & V. Wulf (Eds.), *Social capital and information technology* (pp. 333–354). MIT Press. 13

Belkin, N. J., Oddy, R., & Brooks, H. (1982, June). Ask for information retrieval (part 1). *Journal of Documentation*, *38*(2), 61-71. 6

Berners-Lee, T. (2006, July). Linked data. Available from `http://www.w3.org/DesignIssues/LinkedData.html` 95

BNC. (2007). *British national corpus* (Vol. 3). Oxford University Computing Services on behalf of the BNC Consortium. Available from `http://`

`www.natcorp.ox.ac.uk/` 131

Bojārs, U., & Breslin, J. G. (2005). *Sioc: Semantically-interlinked online communities.* Available from `http://rdfs.org/sioc/spec/` 97

Boje, D. M., & Murnighan, J. K. (1982, October). Group confidence pressures in iterative decisions. *Management Science*, *28*(10), 1187–1196. Available from `http://www.jstor.org/stable/2630947` 102

Bos, N., Olson, J., Gergle, D., Olson, G., & Wright, Z. (2002). Effects of four computer-mediated communications channels on trust development. In *Chi '02: Proceedings of the sigchi conference on human factors in computing systems* (pp. 135–140). New York, NY, USA: ACM. Available from `http://doi.acm.org/10.1145/503376.503401` 62

boyd, d. (2002). *Faceted id/entity: Managing representation in a digital world.* Unpublished master's thesis, Massachusetts Institute of Technology. 192

boyd, d. (2008). *Taken out of context: American teen sociality in networked publics.* Unpublished doctoral dissertation, University of California, Berkeley. 45, 48, 192

Brickley, D., & Miller, L. (2000). Foaf: Friend of a friend. Available from `http://www.foaf-project.org/` 97

Brin, S., & Page, L. (1998, 4). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, *30*(1-7), 107–117. Available from `http://www.sciencedirect.com/science/article/B6TYT-3WRC342-2N/2/63e7d8fb6a64027a0c15e6ae3e402889` 20, 63

Briscoe, B., Odlyzko, A., & Tilly, B. (2006, July). Metcalfe's law is wrong. *IEEE Spectrum*. Available from `http://www.spectrum.ieee.org/computing/networks/metcalfes-law-is-wrong/0` 39

Callahan, S. (2006a, June). Techniques for expertise location. *Anecdote*. Available from `http://www.anecdote.com.au/whitepapers.php?wpid=13` 16

Callahan, S. (2006b, April). Want to manage tacit knowledge? communities of practice offer a versatile solution. *Anecdote*. 12

Cameron, K. (2005a). *The laws of identity.* Available from `http://msdn.microsoft.com/en-us/library/ms996456.aspx` 46

Cameron, K. (2005b). *Microsoft's vision for an identity metasystem.* Available from `http://msdn.microsoft.com/en-us/library/ms996422.aspx` 79

Campbell, C. S., Maglio, P. P., Cozzi, A., & Dom, B. (2003). Expertise identification using email communications. In *Proceedings of the twelfth international conference on information and knowledge management* (pp. 528–531). Available from `http://portal.acm.org/citation.cfm?id=956863.956965` 15

Chase, W. G., & Simon, H. A. (1973, January). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. Available from `http://www.sciencedirect.com/science/article/B6WCR-4D6RJR4-53/2/d368b90478edf5d8b9ee81a894eff330` 66

Choo, C. W. (1996, October). The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. *International Journal of Information Management*, *16*(5), 329–340. 5, 6, 18, 28

Claridge, T. (2004). *Definitions of social capital.* Available from `http://www.gnudung.com/literature/definition.html` 54

Clark, P. B., & Wilson, J. Q. (1961, September). Incentive systems: A theory of organizations. *Administrative Science Quarterly*, *6*(2), 129–166. Available from `http://www.jstor.org/stable/2390752` 33

Collier, P. (2002). Social capital and poverty: a microeconomic perspective. In C. Grootaert & T. V. Bastelaer (Eds.), *The role of social capital in development* (pp. 19–41). Melbourne: Cambridge University Press. 54

Collins, H. M., & Evans, R. (2007). *Rethinking expertise.* Chicago : University of Chicago Press. Available from `http://search.lib.unc.edu?R=UNCb5301057` 65, 69, 70, 72

Collins, H. M., & Pinch, T. (1993). *The golem : What everyone should know about science.* Cambridge University Press. 70

Craswell, N., Hawking, D., Vercoustre, A.-M., & Wilkins, P. (2001). P@noptic expert: Searching for experts not just for documents. In *Ausweb 2001.* 15

Cross, R., & Borgatti, S. P. (2004). The ties that share: Relational characteristics that facilitate information seeking. *Social capital and information technology*, 137–161. 53

Cross, R., & Sproull, L. (2004, July-August). More than an answer: Information relationships for actionable knowledge. *Organization Science*, *15*(4), 446-462. Available from `http://www.jstor.org/stable/30034748` 6

Dalkey, N. C., & Helmer, O. (1963, April). An experimental application of the delphi method to the use of experts. *Management Science*, *9*(3), 458–467. Available from `http://www.jstor.org/stable/2627117` 101

Davenport, T. H. (1997). If only hp knew what hp knows. *Ernst & Young Center for Business Innovation*, 20–25. 16

Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. Unpublished doctoral dissertation, Massachusetts Institute of Technology. 122, 194

de Groot, A. D. (1965). *Thought and choice in chess*. Basic Books. 66

de Groot, A. D. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving: Research, method, and theory* (pp. 19–50). Wiley. 66

Dekker, P., & Uslaner, E. M. (2001). *Social capital and participation in everyday life*. London: Routledge. Available from `http://www.gnudung.com/literature/definition.html` 54

Dervin, B. (1983). *An overview of sense-making research: Concepts, methods, and results to date*. Dallas. Available from `http://www.ideals.uiuc.edu/bitstream/handle/2142/2281/Dervin83a.htm` 3

Dieng, R., Corby, O., Giboin, A., & Ribiére, M. (1999). Methods and tools for corporate knowledge management. *International Journal of Human-Computer Studies*, *51*(3), 567–598. Available from `doi:10.1006/ijhc.1999.0281` 12

DiSo Project. (2007). Diso project. Available from `http://diso-project.org/` 98

Downs, A. (1957). *An economic theory of democracy*. Harper and Row. 2

Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., & Tomkins, A. (2007, August). Visualizing tags over time. *ACM Transactions on the Web*, *1*(2), 7:1–22. 92

Dubner, S. J. (2008, March). How did a-rod get so good? *The New York Times*. Available from `http://freakonomics.blogs.nytimes.com/2008/03/11/how-did-a-rod-get-so-good/` 68

Dubner, S. J., & Levitt, S. D. (2006, May). A star is made. *The New York Times*. Available from `http://www.nytimes.com/2006/05/07/magazine/07wwln_freak.html?ex=1304654400&en=2cf57fe91bdd490f&ei=5090&partner=rssuserland&emc=rss&pagewanted=all` 68

Duguid, P. (2005, April-June). "the art of knowing": Social and tacit dimensions of knowledge and the limits of the community of practice. *The Information Society*, *21*(2), 109-118. 12, 32

Dunbar, R. I. M. (2004a, June). Gossip in evolutionary perspective. *Review of General Psychology*, *8*(2), 100–110. Available from `http://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=gpr-8-2-100&site=ehost-live` 56

Dunbar, R. I. M. (2004b). *The human story*. London: Faber and Faber. 56

Easley, D., & Kleinberg, J. M. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press. 38

Egan, M. E., & Shera, J. H. (1952, April). Foundations of a theory of bibliography. *The Library Quarterly*, *22*(2), 125–137. 73

Ehrlich, K. (2003). Locating expertise: Design issues for an expertise locator system. In *Sharing expertise: Beyond knowledge management.* MIT Press. 12, 14

Engelbart, D. C. (2004). Augmenting society's collective iq. *Keynote, Hypertext 2004*. Available from `http://www.dougengelbart.org/pubs/augment-133319.html` 10

Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (2006). *The cambridge handbook of expertise and expert performance*. Cambridge University Press. 64, 68

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. 68

Etzioni, A., & Bhat, R. (2009). Second chances, social forgiveness, and the internet. *The American Scholar*, *Spring*. Available from `http://www.theamericanscholar.org/?p=1926` 59

Facebook. (2009). Facebook.com. Available from `http://www.facebook.com` 98

Farrell, S., & Lau, T. (2006). Fringe contacts: People-tagging for the enterprise. In *Www '06: Proceedings of the 15th international conference on world wide web* (Vol. Collaborative Web Tagging Workshop). 16

Fellbaum, C. (1998). *Wordnet: An electronic lexical database.* Bradford Books. 131

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. 58

Fidel, R., & Green, M. (2004). The many faces of accessibility: engineers' perception of information sources. *Information Processing & Management*, *40*(3), 563–581. Available from `http://www.sciencedirect.com/science/article/B6VC8-4893TCR-1/2/301534c852f30f0dac58b1db3423a02b` 52

Fisman, R., & Khanna, T. (1999). Is trust a historical residue? information flows and trust levels. *Journal of Economic, Behavior & Organization*, *38*, 79-92. 61

Fitzpatrick, R. B. (1999, September). The community of science, inc. *Medical Reference Services Quarterly*, *18*(3), 57–63. Available from `http://dx.doi.org/10.1300/J115v18n03_04` 14, 16

Fitzpatrick, R. B. (2001, January). The community of science, inc., part 2. *Medical Reference Services Quarterly*, *18*(4), 33–38. Available from `http://dx.doi.org/10.1300/J115v18n04_04` 13, 16

Fleener, J. W., & Prince, J. M. (1997). *Using 360-degree feedback in organizations: An annotated bibliography.* Greensboro, NC: Center for Creative Leadership. 56

Fricker, M. (1998). Rational authority and social power: Towards a truly social epistemology. *Proceedings of the Aristotelian Society*, *19*(2), 159–177. 74

Friedman, B., Peter H. Kahn, J., & Howe, D. C. (2000, December). Trust online. *Communications of the ACM*, *43*(12), 34-40. 61

Friedman, E., & Resnick, P. (2001). The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy*, *10*(2), 173-199. 44

Friedman, T. (2005). *The world is flat: A brief history of the twenty-first century*. Farrar, Straus and Giroux. 86

Gerard, H. B., & Hoyt, M. F. (1974). Distinctiveness of social categorization and attitude toward ingroup members. *Journal of Personality and Social Psychology*, *29*(6), 836–842. Available from `http://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=psp-29-6-836&site=ehost-live` 30

Gerard, H. B., & Orive, R. (1987). The dynamics of opinion formation. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 171–202). Available from `http://dx.doi.org/10.1016/S0065-2601(08)60414-1` 58

Gilder, G. (1993, September). Metcalfe's law and legacy. *Forbes ASAP*. Available from `http://www.seas.upenn.edu/~gaj1/metgg.html` 39

Gilovich, T. (1987). Secondhand information and social judgement. *Journal of Experimental Social Psychology*, *23*, 59–74. 4

Gladwell, M. (2008). *Outliers: The story of success*. Little, Brown and Company. 64

Godin, S. (2008). 10,000 hours. Available from `http://sethgodin.typepad.com/seths_blog/2008/12/10000-hours.html` 69

Goffman, E. (1959). *The presentation of self in everyday life*. Anchor Books. 42, 43, 191

Golbeck, J., & Hendler, J. (2004a). Accuracy of metrics for inferring trust and reputation in semantic web-based social networks. In *Engineering knowledge in the age of the semantic web* (Vol. 3257, p. 116-131). Springer Berlin / Heidelberg. 62

Golbeck, J., & Hendler, J. (2004b). Inferring reputation on the semantic web. In *Proceedings of the 13th international www conference (www 2004)*. 62

Golbeck, J., & Hendler, J. (2004c). Reputation network analysis for email filtering. *Conference on Email and Anti-Spam (CEAS-04)*. Available from `http://citeseer.ist.psu.edu/687785.html` 62

Golder, S. A., & Huberman, B. A. (2005). *The structure of collaborative tagging systems* (Tech. Rep.). Information Dynamics Laboratory, HP Labs. 22, 91, 92

Goldman, A. (2001). Social epistemology. *Stanford Encylopedia of Philosophy*. Available from `http://plato.stanford.edu/entries/epistemology -social/` 74

Granovetter, M. S. (1973, May). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380. 40, 206

Gruber, T. (2005, October). *Tagontology - a way to agree on the semantics of tagging data.* Available from `http://tomgruber.org/writing/ tagontology.htm` 96

Gruber, T. (2008). Collective knowledge systems: Where the social web meets the semantic web. *J. Web Sem.*, *6*(1), 4-13. 37

Gyöngyi, Z., Garcia-Molina, H., & Pedersen, J. (2004). Combating web spam with trustrank. In *Vldb '04: Proceedings of the thirtieth international conference on very large data bases* (pp. 576–587). VLDB Endowment. Available from `http://portal.acm.org/citation.cfm?id= 1316689.1316740` 63

Halpern, D., & Nilan, M. S. (1988, October). A step toward shifting the research emphasis in information science from the system to the user: An empirical investigation of source-evaluation behavior, information seeking and use. *Proceedings of the American Society of Information Science*. 53

Hardin, G. (1968, December). The tragedy of the commons. *Science*, *162*(3859), 1243–1248. Available from `http://www.sciencemag.org/ cgi/content/full/162/3859/1243` 34

Hawking, S. (1988). *A brief history of time.* Bantam Books. 53

Helmer, O., & Rescher, N. (1959). On the epistemology of the inexact sciences. *Management Science*, *6*(1), 25–52. Available from `http://www.jstor.org/stable/2627474` 101

Herling, R. W. (2000, February). Operational definitions of expertise and competence. *Advances in Developing Human Resources*, *2*(1), 8–21. Available from `http://adh.sagepub.com/cgi/content/abstract/2/1/8` 64

Heylighen, F. (1999, October). Collective intelligence and its implementation on the web: Algorithms to develop a collective mental map. *Computational & Mathematical Organization Theory*, *5*(3), 253–280. Available from `http://dx.doi.org/10.1023/A:1009690407292` 35

Hollan, J., Hutchins, E., & Kirsh, D. (2000, June). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, *7*(2), 174–196. 35

Hsu, C.-C., & Sandford, B. A. (2007, August). The delphi technique: Making sense of consensus. *Practical Assessment, Research & Evaluation*, *12*(10). 102

IAU. (2006, August). *Iau0603: Iau 2006 general assembly: Result of the iau resolution votes* (Tech. Rep.). International Astronomical Union (IAU). Available from `http://www.iau.org/public_press/news/detail/iau0603/` 73

Johnson, B., Lorenz, E., & Lundvall, B.-Å. (2002). Why all this fuss about codified and tacit knowledge? *Industrial and Corporate Change*, *11*(2), 245–262. 26

Jones, T. (2009). *Fluidinfo.* Available from `http://fluidinfo.com/` 87

Jøsang, A. (2009). *Subjective logic.* Draft. 80, 81, 82, 83

Jøsang, A., Hayward, R., & Pope, S. (2006). Trust network analysis with subjective logic. *Proceedings of the 29th Australasian Computer Science Conference*, *48*, 85–94. Available from `http://portal.acm.org/citation.cfm?id=1151710` 63

Kalyuga, S. (2009). *Managing cognitive load in adaptive multimedia learning.* Hershey, PA: Information Science Reference. Available from `http://`

search.lib.unc.edu?R=UNCb5805005 66

Kamvar, S. D., Schlosser, M. T., & Garcia-Molina, H. (2003). The eigentrust algorithm for reputation management in p2p networks. In *Www '03: Proceedings of the 12th international conference on world wide web* (pp. 640–651). New York, NY, USA: ACM. Available from `http://doi.acm.org/10.1145/775152.775242` 63

Kelton, K., Fleishmann, K. R., & Wallace, W. A. (2008). Trust in digital information. *Journal of the American Society for Information Science and Technology*, *59*(3), 363-374. Available from `http://dx.doi.org/10.1002/asi.20722` 61

Kent, S. (1994). Words of estimative probability. In D. P. Steury (Ed.), *Sherman kent and the board of national estimates: Collected essays.* Available from `https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/sherman-kent-and-the-board-of-national-estimates-collected-essays/6words.html` 81

Kim, E. E. (2009, March). *The squirm test.* Available from `http://blueoxen.com/blog/2009/03/the-squirm-test/` 58

Kim, H. L., & Breslin, J. G. (2008). *Scot: Social semantic cloud of tags.* Available from `http://scot-project.org/scot/` 97

Kleinberg, J. M. (1999a). Authoritative sources in a hyperlinked environment. *J. ACM*, *46*(5), 604–632. Available from `http://doi.acm.org/10.1145/324133.324140` 20, 63

Kleinberg, J. M. (1999b). Hubs, authorities, and communities. *ACM Comput. Surv.*, *31*(4es), 5. Available from `http://doi.acm.org/10.1145/345966.345982` 40

Knight, I. (2009, November). I'm belle de jour. *The Sunday Times*. Available from `http://entertainment.timesonline.co.uk/tol/arts_and_entertainment/books/article6917495.ece` 49

Knoke, D. (1988, June). Incentives in collective action organizations. *American Sociological Review*, *53*(3), 311–329. 34

Kogut, B., & Zander, U. (1996). What firms do? coordination, identity, and

learning. *Organization Science*, *7*(5), 502–518. 55

Krishnamurthy, B., & Wills, C. (2009). Privacy diffusion on the web: a longitudinal perspective. In *Www '09: Proceedings of the 18th international conference on world wide web* (pp. 541–550). New York, NY, USA: ACM. Available from `http://doi.acm.org/10.1145/1526709.1526782` 49

Kühn, O., & Abecker, A. (1997, August). Corporate memories for knowledge management in industrial practice: Prospects and challenges. *Journal of Universal Computer Science*, *3*(8), 929–954. 13

Lamont, J. (2003, January). Expertise location and the learning organization. *KMWorld*. Available from `http://www.kmworld.com/Articles/PrintArticle.aspx?ArticleID=9410` 11, 13

Lanier, J. (2006, May). Digital maoism: The hazards of the new online collectivism. *Edge*. Available from `http://www.edge.org/3rd_culture/lanier06/lanier06_index.html` 78

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts.* Princeton University Press. 22, 74

Lauterbach, D., Truong, H., Shah, T., & Adamic, L. (2009, August). Surfing a web of trust: Reputation and reciprocity on couchsurfing.com. In *International symposium on social intelligence and networking (sin-09).* Vancouver, Canada. 63

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge University Press. Available from `http://www.amazon.com/exec/obidos/tg/detail/-/0521423740` 12

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., et al. (2003, December). A brief history of the internet. , *3.32*. Available from `http://www.isoc.org/internet/history/brief.shtml` 95

Lessig, L. (1999). *Code and other laws of cyberspace.* New York, NY, USA: Basic Books. 77

Levien, R. (2004). *Attack resistant trust metrics.* Unpublished doctoral dissertation, UC Berkeley. 63

Levitt, S. D., & Dubner, S. J. (2005). *Freakonomics: A rogue economist*

explores the hidden side of everything. William Morrow and Company. 68

Lévy, P. (1997). *Collective intelligence: Mankind's emerging world in cyberspace.* New York: Plenum Press. 35

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140), 1–55. 122

Lin, N. (1999). Building a network theory of social capital. *INSNA Connections*, *22*(1), 28–51. 40, 55

Lin, N. (2001). *Social capital: A theory of social structure and action.* New York: Cambridge University Press. 55

Ling, C. W., Sandhu, M. S., & Jain, K. K. (2009). Knowledge sharing in an american multinational company based in malaysia. *Journal of Workplace Learning*, *21*(2), 125–142. 12

Linstone, H. A., & Turoff, M. (1975). *The delphi method: Techniques and applications.* London: Addison-Wesley. Available from `http://www.is.njit.edu/pubs/delphibook/` 102, 103

Liu, X.-Y., Zhou, Y.-M., & Zheng, R.-S. (2008, July). Measuring semantic similarity within sentences. In *Proceedings of the 7th international conference on machine learning and cybernetics* (pp. 2558–2562). Kunming. 130

Lowenstein, M. (2009, June). Offline social word of mouth influence on brand decision-making more frequent and more powerful than online social media. *The Harris Poll, Harris Interactive*. Available from `http://www.harrisinteractive.com/harris_poll/pubs/Harris_Poll_2009_06_15.pdf` 11

Lueg, C. (2002, February). Management and information technology: Relationship and perspectives. *Upgrade*, *3*(1), 4–7. 25

Luft, J. (1961). The johari window: A graphic model of awareness in interpersonal relations. *NTL Human Relations Training News*, *5*(1), 6–7. 57

Luft, J., & Ingham, H. (1955). The johari window: A graphic model of interpersonal awareness. *Proceedings of the Western Training Laboratory*

in Group Development. 57

Lukas, A. (2008, May). *Mine! paper.* Available from `http://docs.google`
`.com/Doc?id=dgc23h2k_397cgqg3xgh` 98

Luo, L., & Wildemuth, B. M. (2009). Delphi studies. In B. M. Wildemuth
(Ed.), *Applications of social research methods to questions in information*
and library science (chap. 10). Westport, CT: Libraries Unlimited. 103

Lupia, A. (1994, March). Shortcuts versus encyclopedias: Information and
voting behavior in california insurance reform elections. *The American*
Political Science Review, *88*(1), 63–76. 2

Lutters, W. G., Ackerman, M. S., Boster, J., & McDonald, D. W. (2000).
Mapping knowledge networks in organizations: Creating a knowledge
mapping instrument. In *Americas conference on information systems.*
14

Lux, M., Granitzer, M., & Kern, R. (2007, September). Aspects of broad
folksonomies. *Text Information Retrieval Workshop 2007*. Avail-
able from `http://www.slideshare.net/dermotte/aspects-of-broad`
`-folksonomies` 85

Malone, T. W., Laubacher, R., & Dellarocas, C. (2009, February). Harnessing
crowds: Mapping the genome of collective intelligence. *MIT Center for*
Collective Intelligence Working Paper 2009-001. 36

Marchionini, G. (2009). Information concepts: From books to cyberspace
identities. Morgan & Claypool Publishers. 192

Marlow, C., Naaman, M., boyd, d., & Davis, M. (2006). Ht06, tagging paper,
taxonomy, flickr, academic article, toread. *Hypertext 2006*. 87

Martin, B. (2008). Knowledge management. *Annual Review of Information*
Science and Technology, *42*. 7

Marton, C., & Choo, C. W. (2002). A question of quality: The effect of source
quality on information seeking by women in it professions. In E. Toms
(Ed.), *Proceedings of the 65th annual meeting of the american society for*
information science (Vol. 39). Medford, NJ: Information Today. 52

Mathes, A. (2004). *Folksonomies - cooperative classification and*

communication through shared metadata. Unpublished master's thesis, University of Illinois Urbana-Champaign. Retrieved May 1, 2005, from `http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html` 84

Maybury, M., D'Amore, R., & House, D. (2002). Awareness of organizational expertise. *International Journal of Human-Computer Interaction*, *14*(2), 199-217. 65

McDonald, D. W. (2001). Evaluating expertise recommendations. In *Proceedings of the acm 2001 international conference on supporting group work (group)* (pp. 214–223). Available from `http://doi.acm.org/10.1145/500286.500319` 14

McDonald, D. W., & Ackerman, M. S. (1998). Just talk to me: A field study of expertise location. In *Computer supported cooperative work* (p. 315-324). 14

McLuhan, M. (1964). *Understanding media: The extensions of man.* McGraw-Hill. 64

McMillan, D. W., & Chavis, D. M. (1986, January). Sense of community: A definition and theory. *Journal of Community Psychology*, *14*(1), 6–23. 31

McNamara, B. (1971, December). The indian medicine show. *Educational Theatre Journal*, *23*(4), 431-445. 43

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001, August). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444. Available from `http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.soc.27.1.415` 29

Mihalcea, R., Corley, C., & Strapparava, C. (2006). Corpus-based and knowledge-based measures of text semantic similarity. In *Proceedings of the national conference on artificial intelligence* (Vol. 21, pp. 775–780). 130, 205

Miles, A., & Bechhofer, S. (2009). *Skos: Simple knowledge organization system.* Available from `http://www.w3.org/TR/skos-primer/` 97

Milgram, S. (1967). The small world problem. *Psychology Today*, *2*(1), 60–67.

40, 206

Mitnick, K., & Simon, W. L. (2001). *The art of deception: Controlling the human element of security.* John Wiley & Sons, Inc. 99

Moore, G. C., & Benbasat, I. (1991, September). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, *2*(3), 192–222. 122, 124, 194

Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *The Academy of Management Review*, *23*(2), 242–266. Available from `http://www.jstor.org/stable/259373` 55

Nardi, B. A. (1995). Activity theory and human-computer interaction. In *Context and consciousness* (pp. 1–20). MIT Press. 36

Negroponte, N. (1995). *Being digital.* New York: Vintage Books. 45

Nelson, L. H. (1993). Epistemological communities. In L. Alcoff & E. Potter (Eds.), *Feminist epistemologies.* New York: Routledge. 22, 74

Newell, A. (1981). The knowledge level. *Artificial Intelligence*. 25

Nilan, M. S., Peek, R. P., & Snyder, H. W. (1988, October). A methodology for tapping user evaluation behaviors: An exploration of users' strategy, source and information evaluating. *Proceedings of the American Society of Information Science.* 53

Nonaka, I. (1991). The knowledge-creating company. *Harvard Business Review*, *69*(6), p96 - 104. Available from `http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=9201061306&site=ehost-live` 3, 18, 27

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, *5*(1), 14–37. Available from `http://www.jstor.org/stable/2635068` 27

Novak, J., Raghavan, P., & Tomkins, A. (2004). Anti-aliasing on the web. In *Www '04: Proceedings of the 13th international conference on world wide web* (pp. 30–39). New York, NY, USA: ACM. Available from `http://doi.acm.org/10.1145/988672.988678` 49

Oates, G. (2008, January). Many hands make light work. *Flickr Blog.*

Available from `http://blog.flickr.net/en/2008/01/16/many-hands-make-light-work/` 95

Odlyzko, A., & Tilly, B. (2005). *A refutation of metcalfe's law and a better estimate for the value of networks and network interconnections.* Available from `http://www.dtc.umn.edu/~odlyzko/doc/metcalfe.pdf` 39

O'Reilly, C. A. (1982, December). Variations in decision makers' use of information sources: The impact of quality and accessibility of information. *The Academy of Management Journal*, *25*(4), 756–771. 53

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action.* Cambridge University Press. 34

Pariser, E. (2011). *The filter bubble.* Penguin Press. 201, 202

Pedersen, T., & Kolhatkar, V. (2009, June). Wordnet::senserelate::allwords - a broad coverage word sense tagger that maximizes semantic relatedness. In *Proceedings of human language technologies: The 2009 annual conference of the north american chapter of the association for computational linguistics, companion volume: Demonstration session* (pp. 17–20). Boulder, CO. 131

Pedersen, T., Patwardhan, S., & Michelizzi, J. (2004). Wordnet::similarity - measuring the relatedness of concepts. In *Proceedings of the 19th national conference on artificial intelligence* (pp. 1024–1025). San Jose, CA. 131

Pesce, M. (2009, June). Sharing power (global edition), keynote. In *Personal democracy forum.* New York. Available from `http://blog.futurestreetconsulting.com/?p=186` 76

Peterson, E. (2006, November). Beneath the metadata: Some philosophical problems with folksonomy. *D-Lib Magazine*, *12*(11). Available from `http://www.dlib.org/dlib/november06/peterson/11peterson.html` 89

Pinelli, T. E., Bishop, A. P., Barclay, R. O., & Kennedy, J. M. (1993). The information-seeking behavior of engineers. In A. Kent & C. M. Hall (Eds.), *Encyclopedia of library & information science* (Vol. 52, pp. 161–201). Marcel Dekker. 52

Polanyi, M. (1962). The republic of science: Its political and economic theory.

Minerva, *1*(1), 54–73. 69, 73

Polanyi, M. (1966). *The tacit dimension.* NY: Doubleday. 28

Popper, K. (1959). *The logic of scientific discovery* (English ed.). Routledge. 75

Raymond, M. (2008, January). My friend flickr: A match made in photo heaven. *Library of Congress Blog*. Available from `http://blogs.loc.gov/loc/2008/01/my-friend-flickr-a-match-made-in-photo-heaven/` 95

Razavi, M. N., & Iverson, L. (2009, May). Improving personal privacy in social systems with people-tagging. In *Group* (p. 11-20). Available from `http://doi.acm.org/10.1145/1531674.1531677` 98

Reed, D. P. (1999). That sneaky exponential - beyond metcalfe's law to the power of community building.
38, 39

Reeves, B., & Nass, C. (1996). *The media equation.* Cambridge University Press. 11

Reichelt, L. (2007, March). *Ambient intimacy.* Available from `http://www.disambiguity.com/ambient-intimacy/` 31

Requena, F. (2003, March). Social capital, satisfaction and quality of life in the workplace. *Social Indicators Research*, *61*(3), 331–360. Available from `http://www.springerlink.com/content/p2573316450u0n10/` 54

Rheingold, H. (2002). *Smart mobs: the next social revolution.* Basic Books. 35

Rice, L. E., & Mitchell, T. R. (1973, March). Structural determinants of individual behavior in organizations. *Administrative Science Quarterly*, *18*(1), 56–70. Available from `http://www.jstor.org/stable/2391927` 58

Rice, R. E., & Aydin, C. (1991, June). Attitudes toward new organizational technology: Network proximity as a mechanism for social information processing. *Administrative Science Quarterly*, *36*(2), 219–244. 31

Rieh, S. Y. (2002). Judgment of information quality and cognitive authority

in the web. *Journal of the American Society for Information Science and Technology*, *53*(2), 145–161. 53

Rogers, E. M. (1962). *Diffusion of innovations.* Illinois: Free Press of Glencoe. 29

Rohrbaugh, J. (1979, August). Improving the quality of group judgment: Social judgment analysis and the delphi technique. *Organizational Behavior and Human Performance*, *24*(1), 73–92. Available from `http://dx.doi.org/10.1016/0030-5073(79)90017-5` 102

Rosedale, P. (2009, May). Talking with philip rosedale about organizational dynamics. Available from `http://blog.jonudell.net/2009/05/18/talking-with-philip-rosedale-about-organizational-dynamics/` 31, 32

Rowe, G., & Wright, G. (1999). The delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, *15*(4), 353–375. Available from `http://www.sciencedirect.com/science/article/B6V92-3Y9RMNX-2/2/096f9ccd99e01d4ad47703b52486cad8` 101, 102

Rowe, G., Wright, G., & McColl, A. (2005, May). Judgment change during delphi-like procedures: The role of majority influence, expertise, and confidence. *Technological Forecasting and Social Change*, *72*(4), 377–399. Available from `http://dx.doi.org/10.1016/j.techfore.2004.03.004` 23, 102

Rumsfeld, D. (2002, February). *Unknown unknowns.* Press Briefing. 65

Russell, T. (2006). Cloudalicious: Watching tag clouds over time. *Joint Conference on Digital Libraries (JCDL)*, 364. Available from `http://dx.doi.org/10.1145/1141753.1141859` 22, 92, 94

Russell, T. (2008a, October). Tag decay: A view into aging folksonomies. *Proceedings of the American Society of Information Science and Technology*, *44*(1), 1–5. Available from `http://dx.doi.org/10.1002/meet.1450440364` 94, 95

Russell, T. (2008b, February). Watching organizational opinion via social tagging. *iConference 2008*. 94

Russell, T., & Stutzman, F. (2007, October). Self-representation of online

identity in collected hyperlinks. *Proceedings of the American Society of Information Science and Technology*, *44*(1), 1–4. 49

Sabater, J., & Sierra, C. (2005, September). Review on computational trust and reputation models. *Artificial Intelligence Review*, *24*(1), 33-60. 50, 195

Sandhaus, E., & Larson, R. (2009, October). First 5,000 tags released to the linked data cloud. *Open Blog - NYTimes.com*. Available from `http://open.blogs.nytimes.com/2009/10/29/first-5000 -tags-released-to-the-linked-data-cloud/` 96

Sanger, L. M. (2009, February). The fate of expertise after wikipedia. *Episteme*, *6*(1), 52-73. 77

Savolainen, R. (2007, April). Media credibility and cognitive authority. the case of seeking orienting information. *Information Research*, *12*(3). Available from `http://InformationR.net/ir/12-3/paper319.html` 5

Schachter, J. (2005, April). *Del.icio.us.* Retrieved May 1, 2005, from `http:// del.icio.us` 84, 87

Scheibe, M., Skutsch, M., & Schofer, J. (1975). Experiments in delphi methodology. In H. A. Linstone & M. Turoff (Eds.), *The delphi method: Techniques and applications* (pp. 257–281). Addison-Wesley. 102

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(5), 484–494. 67

Schraagen, J. M. (1993). How experts solve a novel problem in experimental design. *Cognitive Science*, *17*(2), 285–309. Available from `http://www.sciencedirect.com/science/article/B6W48 -4B4RPM8-17/2/2ba61b34261c9d217c59b7237d70d771` 67

Seifert, C. F., Yukl, G., & McDonald, R. A. (2003). Effects of multisource feedback and a feedback facilitator on the influence of behavior of managers toward subordinates. *Journal of Applied Psychology*, *88*(3), 561–569. Available from `http://dx.doi.org/10.1037/0021-9010.88.3.561` 57

Sharma, N. (2004). *The origin of the "data information knowledge wisdom" hierarchy.* Available from `http://www-personal.si.umich.edu/~nsharma/dikw_origin.htm` 25

Shirky, C. (2005, May). Ontology is overrated. Available from `http://www.shirky.com/writings/ontology_overrated.html` 85, 86

Shirky, C. (2008a). *Here comes everybody: The power of organizing without organizations.* Penguin Press. 35

Shirky, C. (2008b, September). It's not information overload. it's filter failure. *Web 2.0 Expo NY 08*. Available from `http://web2expo.blip.tv/file/1277460/` 205

Simeonov, S. (2006, July). Metcalfe's law: more misunderstood than wrong? 39

Simon, H. A. (1957). *Models of man: Social and rational.* New York: Wiley. 5

Sinha, R. (2005, September). *A cognitive analysis of tagging.* Retrieved 2009-07-01, from `http://rashmisinha.com/2005/09/27/a-cognitive-analysis-of-tagging/` 85

Skenazy, L. (2009). *Free-range kids: Giving our children the freedom we had without going nuts with worry.* Jossey-Bass. 45

Smith, D. H. (1994). Determinants of voluntary association participation and volunteering: A literature review. *Nonprofit and Voluntary Sector Quarterly*, *23*(3), 243–263. 34

Solomon, R. C. (2000). Trusting. In M. Wrathall & J. Malpas (Eds.), *Heidegger, coping, and cognitive science: Essays in honor of hubert l. dreyfus* (Vol. 2, p. 229-244). Cambridge, MA, USA: MIT Press. 61

Solove, D. J. (2007). *The future of reputation: Gossip, rumor, and privacy on the internet.* Yale University Press. Available from `http://ssrn.com/abstract=1019177` 46

Stein, E. W. (1995, February). Organization memory: Review of concepts and recommendations for management. *International Journal of Information Management*, *15*(1), 17–32. Available from `doi:10.1016/0268-4012(94)00003-C` 12

Sternberg, R. J., & Frensch, P. A. (1992). On being an expert: A cost and benefit analysis. In R. R. Hoffman (Ed.), *The psychology of expertise: Cognitive research and empirical ai* (pp. 191–203). New York: Springer-Verlag. 67

Stutzman, F. (2006). An evaluation of identity-sharing behavior in social network communities. *International Digital and Media Arts Journal*, *3*(1). 45

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations.* Doubleday. Hardcover. Available from `http://www.amazon.fr/exec/obidos/ASIN/0385503865/citeulike04-21` 35

Swan, J., Scarbrough, H., & Preston, J. (1999, June). Knowledge management - the next fad to forget people? *Proceedings of 7th European Conference on Information Systems*, *2*, 668–678. 29

Swanson, R. A. (2007). *Analysis for improving performance : tools for diagnosing organizations and documenting workplace expertise.* San Francisco, CA: Berrett-Koehler Publishers. Available from `http://search.lib.unc.edu?R=UNCb5208200` 64

Sweller, J., Merrienboer, J. van, & Paas, F. (1998, 09 07). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. Available from `http://dx.doi.org/10.1023/A:1022193728205` 3

Tajfel, H., & Turner, J. (1986). The social identity theory of intergroup behaviour. *Psychology of Intergroup Relations*, 7–24. 43, 191

Taylor, F. W. (1911). *The principles of scientific management.* 26

Uslaner, E. (2000, December). Social capital and the net. *Communications of the ACM*, *43*(12), 60-64. 55

Van De Van, A. H., & Delbecq, A. L. (1974). The effectiveness of nominal, delphi, and interacting group decision making processes. *The Academy of Management Journal*, *17*(4), 605–621. Available from `http://www.jstor.org/stable/255641` 102

Vancouver, J. B., & Morrison, E. W. (1995). Feedback inquiry: The

effect of source attributes and individual differences. *Organizational Behavior and Human Decision Processes*, *62*(3), 276–285. Available from `http://www.sciencedirect.com/science/article/B6WP2 -45R89BN-3F/2/fce3391bd9cb119f3038017b50c43d73` 52

Vander Wal, T. (2005, January). *Folksonomy explanations.* Retrieved April 30, 2005, from `http://www.vanderwal.net/random/entrysel.php?blog= 1622` 85

Vander Wal, T. (2007, February). Folksonomy coinage and definition. Available from `http://vanderwal.net/folksonomy.html` 84

Vander Wal, T. (2009, January). Optimizing tagging ui for people and search. *Personal InfoCloud*. Available from `http://www.personalinfocloud .com/2009/01/optimizing-tagging-ui-for-people-search.html` 88

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology*, *95*(5), 1202–1216. 59

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003, September). User acceptance of information technology: Toward a unified view. *Management Information Systems Quarterly*, *27*(3), 425–478. 122, 124, 194, 195

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, Ed.). Cambridge, MA: Harvard University Press. 36

Walther, J. B., Heide, B. V. D., Hamel, L. M., & Shulman, H. C. (2009, April). Self-generated versus other-generated statements and impressions in computer-mediated communication: A test of warranting theory using facebook. *Communication Research*, *36*(2), 229–253. Available from `http://crx.sagepub.com/cgi/content/abstract/36/2/229` 22, 23

Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (pp. 529–563). 22

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5–34. 124

Watts, D. J., & Strogatz, S. H. (1998, June). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440-442. Available from `http://search.epnet.com/login.aspx?direct=true&db=afh&an=698663` 40, 206

Wegner, D. M. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). New York: Springer-Verlag. Available from `http://www.wjh.harvard.edu/~wegner/tm.htm` 18

Weinberger, D. (2005, May). Tagging and why it matters. *Berkman Center for Internet & Society*. Available from `http://cyber.law.harvard.edu/publications/2005/Tagging_and_Why_It_Matters` 87, 89

Weinberger, D. (2006, November). Beneath the metadata - a reply. *self-published*. Available from `http://www.hyperorg.com/blogger/mtarchive/beneath_the_metadata_a_reply.html` 90

Weinberger, D. (2008a). *Everything is miscellaneous: The power of the new digital disorder.* Macmillan. 89

Weinberger, D. (2008b, November). Philosophical problems with folksonomies. *self-published*. Available from `http://www.hyperorg.com/blogger/2008/11/30/philosophical-problems-with-folksonomies/` 90

Weinberger, D. (2009, June). Facts and transparency. In *Personal democracy forum.* New York. Available from `http://www.youtube.com/watch?v=o3qSDLF6lU4` 75, 76

Wesch, M. (2008a, June). An anthropological introduction to youtube. *Presented at the Library of Congress*, Retrieved on Oct 23, 2009. Available from `http://www.youtube.com/watch?v=TPAO-lZ4_hU` 48

Wesch, M. (2008b, July). *Context collapse.* Available from `http://mediatedcultures.net/ksudigg/?p=183` 48

Wikipedia. (2003). No original research. Available from `http://en.wikipedia.org/wiki/Wikipedia:No_original_research` 78

Wilson, P. (1983). *Second-hand knowledge: An inquiry into cognitive authority.* Westport, CT: Greenwood Press. 3, 4, 20, 71

Windley, P. J., Tew, K., & Daley, D. (2007, May). A framework for building reputation systems. In *Www2007.* Banff, Canada. 49

Woudstra, L., & Hooff, B. van den. (2008, May). Inside the source selection process: Selection criteria for human information sources. *Information Processing & Management*, *44*(3), 1267–1278. 53

Yahoo! (2005, April). *Flickr.com.* Available from `http://flickr.com` 88

Yamagishi, T., & Kiyonari, T. (2000). The group as the container of generalized reciprocity. *Social Psychology Quarterly*, *63*(2), 116–132. Available from `http://www.jstor.org/stable/2695887` 55

Yamim, D. (1996). Expert finding systems for organizations: Domain analysis and the demoir approach. In *Ecscw 1999 workshop: Beyond knowledge management: Managing expertise* (pp. 276–283). 14

Zhang, J., Tang, J., & Li, J. (2007). Expert finding in a social network. In *Advances in databases: Concepts, systems and applications, dasfaa 2007* (pp. 1066–1069). 15