Categorization or Communication? Audience Design in Labeling, Organizing, and Finding Files

Emilee Rader
School of Information
University of Michigan
ejrader@umich.edu

ABSTRACT
In a field experiment, I apply theory from psychology and communications to find out whether group information management tasks are governed by the same communication processes as face-to-face and mediated conversation. This paper describes results that both replicate previous research, and expand our knowledge about common ground and audience design when communication is mediated by a co-constructed artifact like a file-and-folder hierarchy. Results indicate that when information producers label and organize information, *common ground* and *audience design* interact in important and unexpected ways, and impact the ability of information consumers to find information. Consumers searching for files are helped more when producers believe they share common ground with their imagined audience, independent of whether the producer and consumer actually have anything in common. This research provides insight into what information might be provided to users, and at what time, to better support these social processes.

Author Keywords
common ground, audience design, group information management, file labeling and organizing

ACM Classification Keywords
H.5.3 Information Interfaces and Presentation: Group and Organization Interfaces – web-based interaction, collaborative computing

INTRODUCTION
Consider the following examples of online information sharing and reuse:

- A scientist needs to locate some procedures and results from an experiment conducted by another researcher in his lab.
- A student learning the open-source, command-line statistical computing environment R needs to find out how to calculate the mode of her dataset.
- A new member of a design team needs to review requirements analysis activities that took place before he joined the team.
- An intelligence analyst needs to consult information collected by other agencies to assess a potential threat.

Finding the information one needs in situations like these is not straightforward. The scientist looking through someone else's experiment procedures and data encounters information that may not be fully documented or organized for consumption by others. The student learning R becomes frustrated when her search for the statistical mode function fails; while a function called "mode" exists, it doesn't actually calculate the mode[1]. The new design team member must navigate a vast intranet repository of documents and artifacts generated through the requirements-gathering process, lacking the context necessary to identify what might be useful. And the intelligence analyst must solve an information puzzle, with pieces scattered across agencies having differing priorities and protocols, and using different vocabulary for the same kinds of things.

Despite differences in the specific details, these situations have four things in common. First, an *information consumer* must locate information that someone else—an *information producer*—created and shared online. Second, *sharing* means posting information to a shared blog, contributing to a wiki, or uploading a file to a shared folder; the information is made available online without specifying a particular recipient [14, 18]. Third, the information that is shared is *explicit*: it has already been captured or documented in some external, concrete way. And fourth, when information producers contribute to a group information system they must *package*, or label and organize the information for others to use; said another way, packaging is the work information producers do that enables a future information consumer to locate and make sense of the information. Effective packaging is not easy; it requires that information producers be aware of the knowledge, information needs, expectations and context of future information consumers who might need the information [12].

Common ground is the mutual knowledge, beliefs and assumptions that people share about each other [6]. It is inferred based on joint membership in cultural communities and through shared perceptual experiences, and accumulates

[1] http://tolstoy.newcastle.edu.au/R/e6/help/09/01/2475.html

Submitted for review to CSCW 2010.

via conversation. A person tailors his utterances to his communication partner, or *audience*, based on his assumptions and beliefs about what the other person knows; these assumptions and beliefs are informed by common ground they share [13]. This theoretical framework has been applied to computer mediated communication [6]; however, it is not clear whether these fundamental properties of language use in communication also play a role when someone is labeling and organizing files that will be shared with others. While a group information repository is not a communications system, words (file and folder labels) are chosen to represent the contents of documents, and also to suggest relationships among groups of documents [5]. Users of group information systems are not random pairs of people; in the best case, they share a work context which includes a common vocabulary. Common ground helps people understand each other in conversation; might the same be true when the "communication" is mediated by a group information system?

I argue that group information management is a social activity, governed by the same communication processes as face-to-face and mediated conversation. I apply theory from psychology and communications in a field experiment, and describe results that both replicate previous research, and expand our knowledge about common ground and audience design when communication is mediated by a co-constructed artifact like a file-and-folder hierarchy. Results of the experiment indicate that when information producers label and organize information, common ground and audience design interact in important and unexpected ways, and impact consumers' ability to find information. I will show in this paper that consumers searching for files are helped more when producers believe they share common ground, whether or not they actually do. This research helps us think about what information to provide to users of group information systems, and at what time, to better support these social processes.

RELATED WORK

Studies of group information systems tend to treat them as tools for storage and sharing of information objects and their metadata, not as tools for communication. For example, Voida [18] created the "Sharing Palette" for granting other individuals access to one's personal files that makes information about what was shared with whom more visually explicit. Whalen et al. [19] designed a "File Manager" and "Sharing Console" to help users become more aware of the usage history of their shared files. And, Tang et al.[17] built "LiveWire", a system that is able to detect similarities and differences among individual enterprise knowledge workers' files; they suggested this kind information could be used to help other individuals find information they need. These approaches are similar in that they focus on providing users with more information about the objects in the system—where they are, who is looking at them, how they've been used in the past, etc.

Other studies provide descriptive accounts of the difficulty users have with finding files, and the ways they try to cope. Conventions are spoken or unspoken rules for how people should behave in certain social situations. Sometimes groups try solve the finding problem by creating labeling and organizing conventions; however, users struggle to adhere to them and tend to slip back over time into their individual, idiosyncratic preferences. The "conventions" approach is rarely successful without significant overhead such as incentives or strict enforcement [2, 11, 12].

These examples from the literature represent two competing approaches to the design of group information systems: attempting to legislate and enforce where things will go (conventions), or providing information about where things are and how they are used (object awareness). But there is a potential third approach that is less explored: what if users had more information about each other? When people exchange information through conversation, despite the fact that our use of language is imprecise and flexible, we are still able to understand one another and communicate effectively. Common ground is necessary for this coordination to take place; as a conversation progresses, participants introduce ideas and vocabulary that become part of their common ground, helping them to develop a sense of what others do and do not know. This helps them tailor their utterances to their listeners so they can communicate more effectively.

In a lab experiment, Fussell and Krauss [7] showed that people label things differently for themselves than for an unknown future person. Participants wrote short descriptions of abstract line drawings to help themselves identify the drawings at a later time, or to help someone else identify them. When participants returned weeks later, they used the descriptions to identify the drawings, and were correct 86% of the time with their own descriptions, 60% of the time with descriptions written for others, and 49% of the time with descriptions written by other people for themselves. Participants also had the highest confidence that they had identified the correct shape based on their own descriptions, followed by descriptions written for others, and finally descriptions by others for themselves. Some researchers theorize that interlocutors develop a mental model of what they assume other individuals know and expect, and use this information to effectively tailor their utterances to their audience [13]. These results suggest that communication processes might affect the labeling and organizing choices of information producers using in a group information system, making a social approach to designing group information systems an intriguing possibility.

RESEARCH QUESTIONS AND HYPOTHESES

The goal of this research is to understand how the influence of *common ground* and *audience design* on labeling and organizing choices in group information systems affects finding behavior. To test this, I designed an experiment that allowed me to detect potential performance differences when participants completed search tasks in file-and-folder hierarchies created by others with whom they shared common ground (or not), and tailored for different audiences.

One kind of common ground is *community membership*, shared by people who have characteristics in common but have never directly interacted. For example, two people who have lived in the same city but never met can be said to share community membership common ground [13]. This would become apparent if one happened to bump into the other on the street and ask for directions—they would very quickly assess their partner's familiarity with the area and tailor their utterances accordingly [9].

I conducted an experiment with three categorical independent variables: *Producer*, *Imagined Audience*, and *Consumer*. The *Producer* labeled and organized files into a hierarchy; for the experiment I recruited Producers from two different intellectual communities, such that some participants would share community membership common ground with each other, and some would not. Producers were instructed to tailor hierarchies for a particular *Imagined Audience*; levels of Audience were Self, or someone from the Same or Different community as the Producer. Finally, the *Consumer* searched for files in the a hierarchy created by a Producer; Consumers in the experiment could be from the Same or Different intellectual community as the Producer.

Based on the literature mentioned in the previous section, and relying heavily on the research design and results of Fussell and Krauss [7], I can make the following predictions about search task performance under different combinations of the above independent variables.

Hypothesis 1: When the hierarchy *Producer*, the *Imagined Audience* for whom the hierarchy was tailored, and the *Consumer* are all from the same community, the Consumer will have the LEAST difficulty with finding.

Hypothesis 2: When the hierarchy *Producer* and the *Imagined Audience* for whom the hierarchy was tailored are from the same community, but the *Consumer* is not, the Consumer will have the MOST difficulty with finding.

Hypothesis 3: When the hierarchy *Producer* and *Consumer*, or the *Imagined Audience* and *Consumer* are from different communities, Consumers will have INTERMEDIATE difficulty with finding.

Hypothesis 4: When the *Imagined Audience* is *Self*, Consumers will have the LEAST difficulty if they are from the same community as the *Producer* and the MOST difficulty when they are from different communities.

METHOD

I conducted a two-part field experiment in which participants used a web-based application created specifically for the experiment, designed to closely resemble the familiar file and folder "desktop metaphor" user interface. I call this a field experiment because it was conducted entirely online, at the convenience of the participants. At no time did participants visit a lab or interact directly with the experimenter; all interactions were conducted by email, including incentive payments which were accomplished by sending Amazon.com gift certificates after participants had completed each phase of the experiment.

In the Organzing phase, participants created labels for a set of short text files, and organized them into a file-and-folder

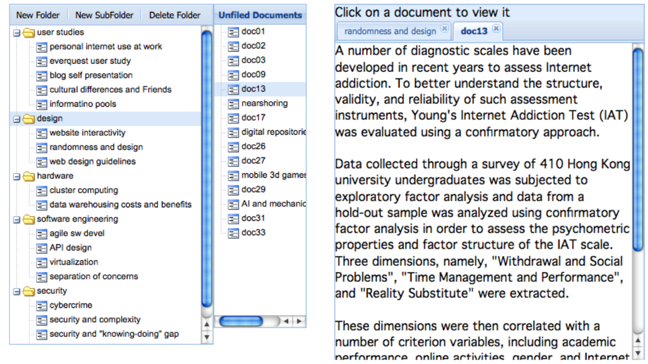

Figure 1: The organizing interface

hierarchy. They were able to view the files online, edit file labels, create and delete folders, and drag and drop files into mutually exclusive folders (i.e., each file could exist in only one place). Figure 1 provides a screen capture of the organizing interface. In the Finding phase, participants later returned to the experiment application and completed a series of search tasks, in which they browsed hierarchies created by other participants to find specific files (Figure 2). The experiment server recorded detailed information about participants' interactions with the system that was later extracted and analyzed. At the end of each phase participants completed a post-questionnaire in which they rated the usability of the system, and answered questions related to the experiment task.

Participants

Participants were 64 graduate students from two departments, Computer Science (CS) and Information Science (IS), at Large Midwestern University. Most CS students were male, and most IS students were female. Participants had been students at the University for one year, on average, with CS students slightly longer than IS students. Sixty-two participants had previous experience sharing files using a group information system. 39% of CS students were non-native speakers compared with 15% of IS students; scores on a verbal ability test consisting of Graduate Record Exam (GRE) practice analogy questions did not differ by community, $F(1,58) = 1.43$, $p = 0.24$.

Text File Selection

The files used in this study were article excerpts, selected from recent issues of online periodicals and trade journals in the summer of 2008. A sample consisting of approximately 50 article excerpts were selected by the experimenter such that they all pertained loosely to current topics in Information and/or Computer Science. Some were more related to IS curriculum, some to CS curriculum, and some potentially interesting to both communities. The excerpts were also chosen to minimize the use of specialized vocabulary wherever possible—the topics were intended to be high-level enough that participants would spend their time and effort in the experiment organizing the files, not attempting to grasp the concepts in each of the texts. Thirty-three files were randomly selected from the sample for use in the experiment.

Table 1: **Organizing conditions, and number of participants**

Imagined Audience	*Producer*: Computer Science	*Producer*: Information Science	(total N)
Same	9	10	(20)
Different	11	11	(21)
Self	11	12	(23)
(total N)	(31)	(33)	64

Labeling and Organizing Procedure

In the Organizing phase, participants first completed a practice session to become familiar with the mechanics of using the interface (see Figure 1). Then, they read instructions that set up the experiment scenario. All were told to imagine they were writing a literature review paper; one-third were instructed to organize the files so they could find them later if they needed to refer back to them. The remaining participants were instructed to imagine themselves collaborating with someone from their own department, or with someone from the opposite department. So, for example, Information Science students were either told to assume they were working with other Information Science students, or students from Computer Science. The instructions participants received were similar to the below:

> On the following screen, you will be presented with a list of files. Each one contains a short article summary or excerpt. You may or may not already be familiar with the topics and concepts in the files. Your task is to create a more descriptive label for each one, and organize them into folders.
>
> There are many different ways to go about completing this task. Some people prefer to read through all of the files and create labels, before organizing them into folders. Others label a few at a time and create folders as they go, renaming and rearranging folders as necessary. What process to follow is completely up to you.
>
> When thinking about what to name the files and what folders to put them in, imagine that you are working on writing a literature review paper for a group project with Information Science graduate students at Large Midwestern University, and other members of your group will need to find some of the files later.
>
> In fact, Information Science students will be invited to participate in Part 2 of this experiment, and they may actually be asked to find files in the hierarchy you will be creating in this part of the experiment. So, please focus on creating a hierarchy with an organizational structure that would make the most sense for Information Science students.

Each participant, or *Producer* constructed a single hierarchy, for one particular *Imagined Audience*. See Table 1 for a depiction of the Organizing conditions, and number of participants in each condition. Participants were told they would not receive the incentive payment if they did not make a good faith attempt to organize the files; three participants were disqualified and replaced when it became apparent that they had not taken the organizing task seriously, from the spurious labels and extremely short time interval to complete the task.

Finding Procedure

About 60 days after organizing the files (mean time interval = 60.02 days, sd = 9.96), 48 participants revisited the experiment system and searched for a sequence of the same files they had organized, each one in a different hierarchy created by another participant.

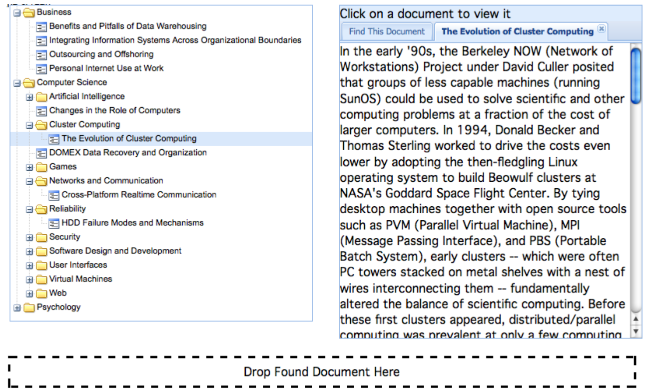

Figure 2: The search tasks interface

Twenty-four participants in the Finding phase were Computer Science (CS) graduate students, and the rest were Information Science graduate students.

Each participant completed two practice search tasks in the Finding interface (see Figure 2), and then 24 experimental search tasks. The experiment system displayed a target file, with no title, and a hierarchy with all of the folders closed; participants browsed the hierarchy, opening and closing folders and viewing files until they found the target. At that point, each participant (or *Consumer*) dragged and dropped the found file into a box at the bottom of the screen, and the next search hierarchy and search target were automatically displayed. Participants were not told who had created the hierarchy, or for what imagined audience. Because participants were able to view the target file for the entire duration of each search task, the finding success rate was nearly 100%. Similar to the Organizing phase, two participants who completed the Finding phase with an excessive number of incorrect searches in an unrealistically short time period (compared with to other participants) were disqualified and replaced with other participants.

Search Task Conditions

This study treats the hierarchies as communication artifacts, conveying information between the person who created it, the *Producer*, and the person searching within it, the *Consumer*. In addition, each *Producer* was instructed to tailor the artifact he or she created for a particular *Imagined Audience*, varied according to the Organizing phase instructions. We can conceptualize the relationships between pairs of these three real and imagined "interlocutors" in terms of the common ground they could potentially share. For example, if the *Producer* and the *Imagined Audience* are both IS graduate students, they are from the **same** community and therefore share some amount of common ground. Likewise, if the *Consumer* is a CS graduate student, they are from a **different** community and do not share much common ground. In another potential combination, a CS student is instructed to create a hierarchy for himself, and a different CS student searches within that hierarchy; here the *Producer* can be considered to share considerable common ground with the *Imagined Audience*, and the *Consumer* has common ground with both.

Audience-Consumer

Producer-Audience	Same	Different
Same	Same | Same (186 searches in 19 hierarchies)	Same | Different (186 searches in 19 hierarchies)
Different	Different | Same (187 searches in 22 hierarchies)	Different | Different (187 searches in 22 hierarchies)
Self	Self | Same (183 searches in 23 hierarchies)	Self | Different (188 searches in 23 hierarchies)

Figure 3: Conditions in the finding phase of the experiment

Figure 3 represents one way to combine pairs of interlocutors into two common ground dimensions by which the search tasks can be categorized: *Producer-Imagined Audience*, and *Imagined Audience-Consumer*. The figure has six cells corresponding to the search task categories. Participants in the Finding phase of the experiment searched for four target files in each of the six search task categories represented in Figure 3, for a total of 24. The files and hierarchy types were presented to half of the participants in one random order, and to the other half in a different random order, to check for potential order effects. Finally, there were 9-12 hierarchies that could potentially be searched-in for each search task category, corresponding to the number of hierarchies created in each condition of the Organizing phase (see Table 1). Finding phase participants completed search tasks in a subset of hierarchies, selected randomly without replacement, from within each category.

ANALYSIS AND RESULTS

As described above, in the Organizing phase of the experiment 64 participants labeled and organized 33 files into file-and-folder hierarchies. Forty-eight participants returned later for the Finding phase and completed a total of 1138 search tasks using the aforementioned hierarchies, created under different *common ground* and *audience design* conditions. The experiment server logged users' actions as they completed the organizing and search tasks, and these logs provided the data from which the measures for the experiment were constructed.

The dependent variable of interest in this experiment is the count of the total number of clicks (total.clicks) required to find the target file in each of the search tasks. Smaller numbers of clicks mean better performance, i.e., participants were able to find the target file more easily, using fewer actions. Twenty-one of the 1138 search tasks yielded total.clicks greater than 50 (mean = 82.67); these values are remarkably extreme given that the mean file depth over all the hierarchies was 2.39 levels (sd=0.54), the mean number of folders was 9.55 (sd=4.05), and the mean folder size was 4.13 files (sd=2.21). These outliers were removed from the analysis, resulting 1117 total observations (20 to 24 per participant).

Analysis of Variance/Covariance is a common approach when analyzing data in which factors have been experimentally manipulated. However, I wanted to achieve four goals through this analysis, some of which are more easily accomplished using a generalized linear regression model:

1. Control for participant-, task- and hierarchy-level influences on the dependent variable, separately from the experimentally manipulated factors;
2. Conduct statistical hypothesis tests of the experimentally manipulated common ground and audience design factors;
3. Generate model predictions indicating the size of the differences between experiment conditions after controlling for other sources of variability (see #1);
4. And finally, compare these predictions based on the experiment data against theoretical predictions.

I used the R statistical computing environment[2] to model the data using poisson regression. Poisson regression is more appropriate for count data like total.clicks because the assumption of normality is violated. The poisson regression model was estimated using maximum likelihood estimation, with a log link function and errors distributed according to a negative binomial distribution. The negative binomial distribution was necessary because these data are overdispersed; poisson regression yields standard errors that are too low, as well as a poor model fit. By using the negative binomial distribution, the model estimates an additional "dispersion parameter" and thereby compensates for the increased variability. Using the negative binomial distribution allowed me to perform more conservative statistical significance tests on the model estimates, reducing the probability of making a Type I error [4].

Regression Model

The dependent variable in the model is total.clicks, the total number of clicks (consisting of all folder open, folder close, and file view events) to locate the target file. The regressors are:

- imagined.audience: the *Imagined Audience* for whom the hierarchy was created
- PA.Same: are the *Producer* and *Imagined Audience* from the same community? Yes or No
- AC.Same: are the *Imagined Audience* and *Consumer* from the same community? Yes or No
- imagined.audience * AC.Same: 2-way interaction
- PA.Same * AC.Same: 2-way interaction

The controls included in the model are:

- shortest.path: for each search task, the depth in the hierarchy of the target file, i.e., the absolute minimum number of clicks to find the target

[2] http://www.r-project.org/, using glm.nb from the VR bundle

- $\mathrm{average.path.length}$: the average number of steps from any file in a hierarchy to any other file, used as an indication of the complexity of the hierarchy; for example, a hierarchy with files grouped into only two folders at the same level has a lower $\mathrm{average.path.length}$ than a hierarchy with 4 or 5 levels and fewer files per folder
- $\mathrm{consumer.id}$: because each person experienced all types of search tasks, the model includes a fixed effects control for individual differences

In within-subjects experiment designs like this one, variation due to participant individual differences is often modeled using random rather than fixed effects. With random effects, the participant-level effects are estimated from a distribution based on the observed values for the participants in the experiment, rather than including a dummy variable for each participant. Using fixed effects allows for greater precision and statistical power than random effects, but at the expense of the ability to make statistical inferences. This tradeoff is acceptable here because I do not claim the participants in this experiment are representative of the population of all group information system users, and am not using statistical inference to generalize beyond this set of observations. Rather, this model is a tool that can be used to understand what is going on in this particular dataset; any implications based on the results presented here will come from logical, rather than statistical generalization. Finally, in a "mixed model" that includes both fixed and random effects and a non-normal distribution assumption, it is difficult to determine the appropriate degrees of freedom for the random effects, making statistical hypothesis tests of the regressor estimates impractical [3]. This was one of the stated goals of the analysis, so fixed effects are more appropriate here.

The model is constructed as follows:

$$\begin{aligned}\mathrm{log(total.clicks)} = f(&\mathrm{shortest.path, average.path,}\\ &\mathrm{imagined.audience, PA.Same, AC.Same,}\\ &\mathrm{imagined.audience} * \mathrm{AC.Same,}\\ &\mathrm{PA.Same} * \mathrm{AC.same, consumer.id})\end{aligned}$$

Deciding what regressors to use in a model like this is an iterative process that involves specifying the model with different combinations of predictor variables and comparing the model variations using likelihood ratio (LR) tests. LR tests compare deviance, which is a goodness-of-fit indicator, between different models [1]. I performed two LR tests to compare the goodness of fit between the final model, specified above, and three variations. First, I compared the model above with the "saturated model", a statistical construct that includes one regressor for every observation. As such it is able to perfectly predict the observed values. The LR test against the saturated model tests the null hypothesis that the saturated model and less-specified model are effectively the same. If this test results in a p-value above the threshold for significance, the null hypothesis is retained, and the less-specified model is an adequate fit for the data. For the LR test comparing the above model against the saturated model, the p-value was 0.32, indicating that the less-specified model is a reasonable fit. Similarly, a likelihood ratio test comparing the final model with a model including all possible two- and three-way interactions was not significant (p=0.23).

Finally, there were two predictors that could impact the results of the experiment, but were not included in the model: the length of the time interval between the Organizing phase and the Finding phase, and which random order of hierarchies and search target files participants viewed. These two variables are highly collinear with the $\mathrm{consumer.id}$ regressor, and because $\mathrm{consumer.id}$ is a necessary part of the model unwise (and unnecessary) to include these other collinear regressors. The model estimates are accurate regardless; however, it is not possible for the model to estimate the amount of potential influence of either time interval or random task order on the dependent variable if they are not included as regressors.

Instead, I calculated Pearson correlations to assess whether these variables were related to the total number of clicks to find the search target. The correlation between time interval and $\mathrm{total.clicks}$ is 0.0006 (t = 0.021, df = 1115, p = 0.9832), and the correlation between task order and $\mathrm{total.clicks}$ is -0.009 (t = -0.3076, df = 1115, p = 0.7584). For the sake of comparison: the correlation between $\mathrm{shortest.path}$, one of the control measures included in the model, and $\mathrm{total.clicks}$ is 0.21 (t = 6.9958, df = 1115, $p < 0.000$). These correlations indicate that time interval and task order did not influence the outcome of the experiment.

Results

The regression results are detailed in Table 2. The Regressors are the explanatory variables and controls included in the model. For each regressor, there is an estimated coefficient. Remember that this model uses a log transform of the dependent variable; the model predicts the log of the $\mathrm{total.clicks}$ to find a search target rather than the actual count. The estimated coefficients are in the same units and must be transformed back before they can be easily interpreted.

Consider the estimate for $\mathrm{shortest.path}$. If the shortest path to the target file were to increase by one unit, the difference in the log of the expected number of clicks is predicted to increase by 0.19 units, holding other regressors in the model constant. The estimate represents the log of the ratio of the expected number of clicks when $\mathrm{shortest.path}$ is 0 vs. when it is 1. This is difficult to conceptualize in terms of quantity of impact on a particular search task. Interpretation is made easier by transforming the estimate to represent a percentage change in $\mathrm{total.clicks}$ for every 1-click difference in the shortest path length. Calculating the percentage change is fairly simple: exponentiate the estimate, subtract 1, and multiply by 100. For shortest path, this yields 20.86%. So now we can say that for each 1-click increase in the shortest path to reach the target document, the total number of clicks to find the search target increases by 20.86%.

A Wald test was performed on each estimate to test the null hypothesis that true estimate of the coefficient is zero. The Wald tests allow me to test experimental hypotheses, similar in logic to the F-test in ANOVA. These significance tests are

Table 2: Negative Binomial Regression estimates, IRR, % Change. Theta (dispersion paramater) = 2.728. consumer.id **dummy variable coefficients are not included here.**

	Regressors	*Estimates*	*% Change*	*Std. Error*[3]	
0.	(Intercept)	0.744	(2.10 clicks)	0.244	**
1.	shortest.path	0.189	20.862	0.048	***
2.	average.path.length	0.229	25.740	0.056	***
3.	imagined.audience (Info. Sci.)	−0.012	−1.217	0.130	
4.	imagined.audience (Self)	0.134	14.351	0.114	
5.	PA.Same (Yes)	−0.183	−16.749	0.089	*
6.	AC.Same (Yes)	−0.062	−6.009	0.150	
7.	imagined.audience (Info. Sci.) * AC.Same (Yes)	0.0956	10.035	0.226	
8.	imagined.audience (Self) * AC.Same (Yes)	−0.167	−15.384	0.191	
9.	PA.Same (Yes) * AC.Same (Yes)	0.037	3.771	0.125	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1;

most interesting for the experimentally manipulated factors: imagined.audience, PA.Same, and AC.Same (and the interactions). Table 2 shows that the only significant estimate is PA.Same. The Intercept, and two controls in the model are also significant (shortest.path and average.path.length). The lack of significance of the other estimates does NOT mean those estimates are somehow biased or less reliable; it simply means that in the context of this particular set of regressors included in the model, we cannot statistically conclude that the actual coefficient is different from zero.

Model Interpretation

Transforming the model estimates into percent change is only the first step in interpretation. What does a 16.75% decrease in total.clicks for PA.Same, holding everything else in the model constant, really mean in practice for participants in the experiment? Estimates and percent change values for the regressors must be interpreted in the context of the rest of the model, and that means starting from the Intercept. Because most of the regressors in this model are categorical, the concept of these regressors having a value of zero does not really make sense. So instead of thinking about the Intercept as the value of the dependent variable when all other coefficients are zero, it is actually the total.clicks for a particular combination of the categorical variables selected to be the baseline by the model.

For this model, the base rate of clicks per search task is 2.10 when shortest.path and average.path.length are zero, and the categorical regressors take on the values of *Producer-Audience* (Different), and *Audience-Consumer* (Different). Figure 3 depicts the possible combinations of the categorical regressors. One additional categorical dimension can be layered on top of these: *Imagined Audience* type. The Intercept takes on the value imagined.audience (CS); other levels of this categorical regressor that appear in Table 2 are (IS) and (Self).

Table 3 presents the results of the model, interpreted not as coefficient estimates but as differences from the Intercept, and compared with theoretical predictions based on the literature. The rows in the table correspond to the ten points in the fitted values graph (Figure 4). The Intercept, row 5 in the table, is the baseline against which the percent change values are added or subtracted. The percent change numbers come from combining the appropriate model estimates of the effect of different levels of the categorical regressors. Consider, for example, the model prediction of +2.16% (row 7 in the table). This number is calculated by adding the estimates from Table 2 for imagined.audience (Info. Sci.) in row 3, AC.Same (Yes) in row 6, and imagined.audience (Info. Sci.) * AC.Same (Yes) in row 7, and then transforming them into a percent change. The rest of the percentages in that column can be calculated in a similar manner.

Looking at the fitted values in Figure 4, it is clear that in the context of this experiment, the percentage differences between search task conditions translated into at most a 2-click difference between the highest and lowest points in the graph. This raises a question about statistical vs. practical significance; statistical magic aside, is this difference really large enough to be important? My argument is yes. A real-world group information system would likely be broader and deeper and contain more files than the hierarchies created in this experiment, increasing both the shortest path to the target file and the overall complexity of the hierarchy. This is reflected in the model estimates, translated to percentage change, for shortest.path (21% increase in total.clicks) and average.path.length (26% increase in total.clicks).

In general, Consumers performed best (fewest clicks to find the target file) when the *Producer* created a hierarchy for an *Imagined Audience* from the same community, regardless of the community the *Consumer* was from. Consumers had the the most difficulty when searching in hierarchies created by a *Producer* for an *Imagined Audience* that was not like them. This is an interesting and unexpected result; it means that who the Producers THOUGHT their audience was, turned out to be more important than who the Consumers ACTUALLY were. Said another way, Producers created hierarchies in which everyone could find stuff more easily, regardless of what community they were from, but only when they imagined that they were organizing for somebody like them. When *Producer* == *Imagined Audi-*

[3]The studentized Breusch-Pagan test was significant (B = 91.39, df = 56, p = 0.0018), which indicates that heteroskedasticity is present and standard errors are likely underestimated. Table 2 reports White's robust standard errors [8] which are more conservative in the presence of heteroskedasticity; Wald tests on the estimates reflect the adjusted standard errors, reducing the probability of Type I error.

Table 3: Model Results compared with Theoretical Predictions. Model results are presented as % Change (from the Intercept) in total clicks to find the search target; "Best" means fewest clicks.

	Regression Model Results		$\Rightarrow$	*Theoretical Predictions*	*Producer & Consumer*	*Audience & Consumer*	*Producer & Audience*	*Imagined Audience*
1.	-22.39%	Best	$=$	Best	Same	Same	Same	Self
2.	-18.80%	Best	$=$	Best	Same	Same	Same	Comp. Sci.
3.	-11.74%	Best	$=$	Best	Same	Same	Same	Info. Sci.
4.	-1.22%	Worst	$\downarrow$	Intermediate	Same	Different	Different	Info. Sci.
5.	(Intercept)	Worst	$\downarrow$	Intermediate	Same	Different	Different	Comp. Sci.
6.	-6.01%	Intermediate	$=$	Intermediate	Different	Same	Different	Comp. Sci.
7.	$+2.16\%$	Worst	$\downarrow$	Intermediate	Different	Same	Different	Info. Sci.
8.	-4.80%	Intermediate	$\uparrow$	Worst	Different	Different	Same	Self
9.	-17.76%	Best	$\uparrow$	Worst	Different	Different	Same	Info. Sci.
10.	-16.75%	Best	$\uparrow$	Worst	Different	Different	Same	Comp. Sci.

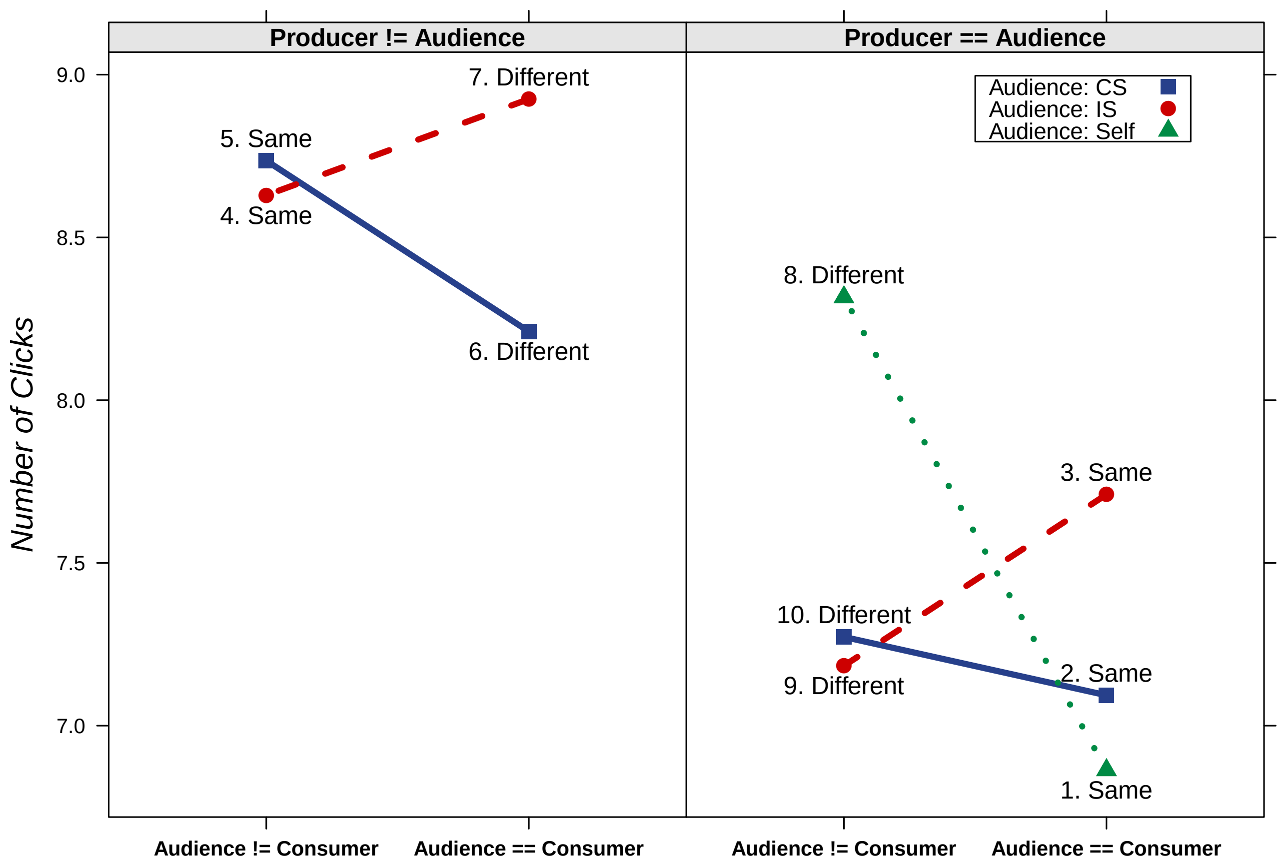

Figure 4: Model results represented as fitted values, based on the median consumer.id **estimate and mean** shortest.path **(2.48 clicks) and mean** average.path.length **(3.93 clicks). "Same" and "Different" point labels refer to *Producer* & *Consumer* community membership, and row numbers in Table 3.**

The lines on the graph illustrate the three *Imagined Audience* conditions: CS, IS, and Self. The left and right panels represent whether or not the *Producer* and his *Imagined Audience* are from the same community. Within each panel, the left and right depict whether the *Imagined Audience* and *Consumer* are from the same community. There are three things to notice about this graph. First, the difference between the left and right panels corresponds to the only significant model estimate for an experimentally manipulated factor (PA.Same). Regardless of the *Consumer* community, participants performed better (fewer clicks) when the *Producer* believed the *Imagined Audience* was similar to them. Second, the model predictions for the Self condition replicate the findings of Fussell and Krauss [7]. Finally, the base rate of clicks to find the search target, for mean shortest.path and mean average.path.length and holding all other factors in the model constant, is 6.41 clicks. (Also note that the y-axis starts around 6.75, not zero.)

ence, all Consumers found the target in fewer clicks, regardless of whether they were like the Producers, or members of the target audience category (see Figure 4).

The model results can be used to evaluate the hypotheses outlined at the beginning of this paper:

Hypothesis 1: When the hierarchy *Producer*, the *Imagined Audience* for whom the hierarchy was tailored, and the *Consumer* are all from the same community, the Consumer will have the LEAST difficulty with finding. This prediction says that the best possible situation for a Consumer is to search in a hierarchy created by another Producer like them, tailored for someone from the same community. In this case, common ground is shared all around, and audience design is easy. This hypothesis was **Confirmed**. Rows 1-3 in Table 3 show that the fewest number of clicks were required in search tasks with these characteristics.

Hypothesis 2: When the hierarchy *Producer* and the *Imagined Audience* for whom the hierarchy was tailored are from the same community, but the *Consumer* is not, the Consumer will have the MOST difficulty with finding. The logic behind this is that both common ground and audience design work against the Consumer, who is from a different community. Surprisingly, this hypothesis was **Rejected.** Rows 8-10 in Table 3 show that where the literature predicted the worst performance, Consumers experienced some of their best performance. This is due to the *Producer-Imagined Audience* effect described above.

Hypothesis 3: When the hierarchy *Producer* and the *Consumer*, or the *Imagined Audience* and *Consumer* are from different communities, Consumers will have INTERMEDIATE difficulty with finding. Under these circumstances, it was expected that despite the *Producer* being instructed to tailor the hierarchy for someone from the opposite community, the common ground and audience design effects would be more important. This hypothesis was also unexpectedly **Rejected**. Rows 4-7 in Table 3 show that Consumers experienced the most difficulty under these search task conditions; the *Producer-Imagined Audience* effect is apparent here as well. All four rows say "Different" in the *Producer* & *Imagined Audience* column.

Hypothesis 4: When the *Imagined Audience* is *Self*, Consumers will have the LEAST difficulty if they are from the same community as the *Producer* and the MOST difficulty when they are from different communities. The prediction says that when a Producer customizes a hierarchy for herself, a Consumer from the same community uses 17% fewer clicks to find the target file than a Consumer from the opposite community (rows 1 and 8 in Table 3). This hypothesis was **Confirmed**, and is a replication of the Fussell and Krauss experiment [7].

DISCUSSION AND FUTURE WORK

This research was conducted with three goals in mind. The first objective was to determine whether communication processes are at work in group information management tasks. Given the results of this study, it is clear that the answer is yes. This suggests thinking about labeling and organizing not just as storage and categorization, but as a communicative activity.

The second goal was to better understand influences of common ground and audience design on hierarchy creation and finding behavior, while replicating previous work. This goal was also accomplished. The main finding of this work is that information Consumers, searching in hierarchies by Producers who organized with similar others in mind, searched most efficiently. In contrast, when Producers tailored their hierarchies to dissimilar others, Consumers required more clicks to reach the search target. It is unlikely that these results came about due to a lack of motivation on the part of Producers who labeled and organized for people not like them. In the post-Organizing questionnaire, participants were asked (using a 5-item Likert scale), "When organizing the documents, how much did you think about what might make sense to [You vs. someone from the Same / Different community] later, if you had to find one of the documents again?" There were no differences in answers to this question by *Imagined Audience* (Kruskal-Wallis chi-squared = 4.5319, df = 2, p = 0.1037) or by *Producer* community (Kruskal-Wallis chi-squared = 0.3938, df = 1, p = 0.5303). Replicating previous findings lends credibility to these results, and provides confidence that what happened in this experiment is indicative of larger patterns rather than local variations.

This raises the question, why might some of these results differ from predictions based on the literature? Consumers underperformed expectations when Producers tailored their hierarchies for different others, and did better than expected when Producers organized for similar others. There are two possible reasons for this. The audience design condition was more nuanced in this experiment than in Fussell and Krauss [7]; rather than just "self" and "other", this experiment had two different flavors of "other" depending on the community membership of the participant. Providing participants with more information to use when doing audience design could easily have allowed for greater nuance in the results. Also, the instructions participants received referred to real communities, and the organizing and labeling task has greater external validity.

The third goal of this research was to gain insight into ways we might design systems that incorporate better support for social aspects of group information management. Given the *Producer-Imagined Audience* result, and in light of previous audience design research, I suspect that finding ways to incorporate support for the formation of more accurate mental models of other users could help. For example, Wittwer et al. [20] found that the level of detail in experts' models of laypersons' knowledge was important for successful communication; mental models that were either oversimplified or too complex proved to be less effective. The same might be happening in this experiment, and would help explain the pattern of results. However, because I did not experimentally manipulate participants' mental models of the audience, making a stronger claim about this is left for future work. Also, this new audience design hypothesis does not fully explain why ALL participants who searched in hierarchies by Producers that organized for similar others did better; this means, for example, that both CS and IS people had an easier time searching in hierarchies created by a CS person who imagined their audience was similar to them. Analysis of data from the Organizing phase of the experiment is currently underway; this will hopefully shed more light on qualitative differences between the types of hierarchies that contributed to these results.

What is most interesting about these results is that a manipulation of experiment instructions was all that was necessary to bring about the observed differences. Producers simply imagined someone similar to them while organizing, and this helped *everyone* in the experiment perform better. There was no special reward offered to experiment participants for making hierarchies that were easier to search, and yet the pattern of results is still unmistakable. As Sen et al. [15] and others [10, 16] have shown, user behavior can be influenced merely by the information designers include in the user interface. These results hint that incorporating information that makes the audience more salient could go a long way toward helping users find the information they need in group information systems.

ACKNOWLEDGMENTS

I would like to thank Rick Wash for many late-night conversations about statistics, and writing code for the experiment system. Also Libby Hemphill and Morgan Daniels for comments on an earlier draft, and Judy Olson for her feedback on the experiment design. Finally, I would like to thank the Rackham Graduate School at UM for funding this research.

REFERENCES

1. A. Agresti. *An Introduction to Categorical Data Analysis*. Wiley, second edition, 2007.
2. L. M. Berlin, R. Jeffries, V. L. O'Day, A. Paepcke, and C. Wharton. Where did you put it? issues in the design and use of a group memory. In *CHI '93*, pages 23–30, Amsterdam, The Netherlands, 1993. ACM Press.
3. B. M. Bolker, M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, 24(3):127–135, 2009.
4. A. L. Byers, H. Allore, T. M. Gill, and P. N. Peduzzi. Application of negative binomial modeling for discrete outcomes: A case study in aging research. *Journal of Clinical Epidemiology*, 56(6):559 – 564, 2003.
5. J. M. Carroll. The role of context in creating names. *Discourse Processes*, 3(1):1–24, 1980.
6. H. H. Clark and S. E. Brennan. Grounding in communication. In L. B. Resnick, J. M. Levine, and S. D. Teasley, editors, *Perspectives on Socially Shared Cognition*, pages 127–149. APA, Washington DC, 1991.
7. S. R. Fussell and R. M. Krauss. The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology*, 25(3):203–219, 1989.
8. D. N. Gujarati. *Basic Econometrics*. McGraw-Hill/Irwin, 4th edition, 2003.
9. E. Isaacs and H. H. Clark. References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116(1):26–37, 1987.
10. C. A. Lampe, N. Ellison, and C. Steinfield. A familiar face(book): profile elements as signals in an online social network. In *CHI '07*, pages 435–444, New York, NY, USA, 2007. ACM.
11. G. Mark and W. Prinz. What happened to our document in the shared workspace? the need for groupware conventions. In *IFIP TC13 Int'l Conf. on HCI*, pages 413–420, 1997.
12. M. L. Markus. Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. *J MIS*, 18(1):57 – 93, 2001.
13. R. S. Nickerson. How we know – and sometimes misjudge – what others know: imputing one's own knowledge to others. *Psychological Bulletin*, 125(6):737–759, 1999.
14. E. Rader. Yours, mine and (not) ours: Social influences on group information repositories. In *CHI '09*, pages 2095–2098, 2009.
15. S. Sen, S. K. Lam, A. M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F. M. Harper, and J. Riedl. tagging, communities, vocabulary, evolution. In *CSCW '06*, pages 181–190, 2006.
16. N. S. Shami, K. Ehrlich, G. Gay, and J. T. Hancock. Making sense of strangers' expertise from signals in digital artifacts. In *CHI '09*, pages 69–78, New York, NY, USA, 2009. ACM.
17. J. C. Tang, C. Drews, M. Smith, F. Wu, A. Sue, and T. Lau. Exploring patterns of social commonality among file directories at work. In *CHI '07*, pages 951–960, New York, NY, USA, 2007. ACM.
18. S. Voida, W. K. Edwards, M. W. Newman, R. E. Grinter, and N. Ducheneaut. Share and share alike: exploring the user interface affordances of file sharing. In *CHI '06*, Montreal, Quebec, Canada, 2006.
19. T. Whalen, E. G. Toms, and J. Blustein. Information displays for managing shared files. In *CHiMiT '08: Proceedings of the 2nd ACM Symposium on Computer Human Interaction for Management of Information Technology*, pages 1–10, New York, NY, USA, 2008. ACM.
20. J. Wittwer, M. Nückles, and A. Renkl. Is underestimation less detrimental than overestimation? the impact of experts' beliefs about a layperson's knowledge on learning and question asking. *Instructional Science*, 36(1):27–52, 2008.