The Effect of Audience Design on Labeling, Organizing, and Finding Shared Files

Emilee Rader
Center for Technology and Social Behavior
Northwestern University
ejrader@northwestern.edu

ABSTRACT

In an online experiment, I apply theory from psychology and communications to find out whether group information management tasks are governed by the same communication processes as conversation. This paper describes results that replicate previous research, and expand our knowledge about *audience design* and packaging for future reuse when communication is mediated by a co-constructed artifact like a file-and-folder hierarchy. Results indicate that it is easier for information consumers to search for files in hierarchies created by information producers who imagine their intended audience to be someone similar to them, independent of whether the producer and consumer actually share common ground. This research helps us better understand packaging choices made by information producers, and the direct implications of those choices for other users of group information systems.

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

General Terms

Human Factors

INTRODUCTION

Consider the following examples of online information sharing and reuse:

- A scientist needs to locate 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 or uploading content to a shared online system; the information is made available online without specifying a particular recipient [13, 17]. 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 inherently social; it requires that information producers be aware of the knowledge, information needs, expectations and context of future information consumers who might need to find the information [9].

People engage in *audience design* during conversation, tailoring their utterances to their communication partner based on assumptions and beliefs about what the other person knows. These assumptions and beliefs change over time through repeated interactions, and are informed by *common ground*

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

CHI 2010, April 10 – 15, 2010, Atlanta, Georgia, USA

[10]. *Common ground* is the mutual knowledge, beliefs and assumptions that people share about each other [5]. It is inferred based on joint membership in cultural communities and through shared perceptual experiences, and accumulates via conversation. This theoretical framework has been previously used to explain behavior in computer mediated communication [5]; it may also apply 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 files, and also to suggest relationships among groups of files [4].

However, it is not clear how *audience design* affects packaging information for reuse, or what the impact on information finding might be, especially when the "audience" is implicit or ill-defined. How do users' beliefs about who will be using a group information system in the future affect labeling and organizing of information contributed to the system? And what effect, if any, does this have for whether people can find the information they need? In this paper, I apply theory from psychology and communications in an online experiment, and describe results that both replicate previous research, and expand our knowledge about packaging and *audience design* when communication is mediated by a co-constructed artifact like a file-and-folder hierarchy.

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 [17] 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. [18] 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.[15] built "LiveWire", a system that is able to detect similarities and differences among individual enterprise knowledge workers' files; they suggested this type of 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. Sometimes groups try to make finding files easier by creating explicit rules for how files should be labeled and organized; however, this approach is rarely successful without significant overhead such as incentives or strict enforcement. Users struggle to adhere to the rules, and slip back over time into their individual, idiosyncratic preferences [1, 8, 9].

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 (rules), 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 via 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 [5].

In a lab experiment, Fussell and Krauss [6] showed that people engage in *audience design* when labeling. 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 others know and expect, and use this model when tailoring their utterances to their audience [10].

I focus on *audience design* as a social process with important consequences for sharing information online. At a high level, I hypothesize that information producers make choices about how files should be labeled and organized that are influenced by their knowledge, beliefs, and assumptions about other users. These choices determine how the information in the system is structured, and the information structure affects whether or not consumers can find what they need.

RESEARCH QUESTIONS AND HYPOTHESES

The goal of this research is to understand how the influence of *audience design* on labeling and organizing choices in group information systems affects finding behavior. To test this, I conducted an experiment that allowed me to detect performance differences when participants completed search tasks in file-and-folder hierarchies. The experiment had 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. *Community membership* common ground is shared by people who have characteristics in common but have never directly interacted; for example, two people who live in the same city but have never met can be said to share community membership common ground [10]. Producers were instructed to tailor the hierarchy they created for a particular *Imagined Audience*; levels of Audience were Self, or someone from the Same or Different intellectual community as the *Producer*. Finally, the *Consumer* searched for files in hierarchies created by an

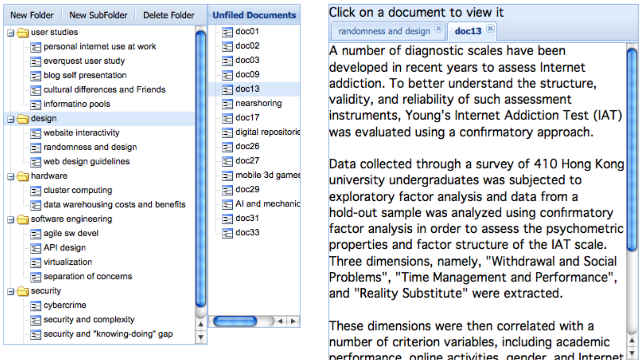

Figure 1: The organizing interface

Producer; Consumers in the experiment could be from the Same or Different intellectual community as the *Producer* or the *Imagined Audience*.

The predictions below follow from two assumptions based on the body of literature mentioned in the previous section: 1) audience design is beneficial in that it helps people achieve shared understanding; and 2) participants from the same intellectual community have more in common than participants who are from different communities, and can therefore communicate more effectively.

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 online 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. The experiment was entirely web-based, conducted 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.

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

In the Organizing phase, participants created labels for a set of short text files, and organized them into a file-and-folder 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 presented to them in the interface (Figure 2).

Participants

Participants were 64 graduate students from two academic units, Computer Science (CS) and the School of Information (IS), at the University of Michigan. Information science and computer science students typically attend courses on different campuses, and have few opportunities for interaction. These two communities are different, and yet similar enough to imagine members actually sharing files in the real world, perhaps on an interdisciplinary team. Most CS students were male, and most IS students were female. 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 group, $F(1,58) = 1.43$, $p = 0.24$. Participants within a community agreed with each other on the same words for a given file about 22% of the time, on average (after stemming and stop word removal). However, across communities, their agreement was around 12%.

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 was selected by the experimenter such that all articles 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 relevant to both communities. The excerpts were 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 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.

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. 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. 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 other participants) were disqualified and replaced with other participants.

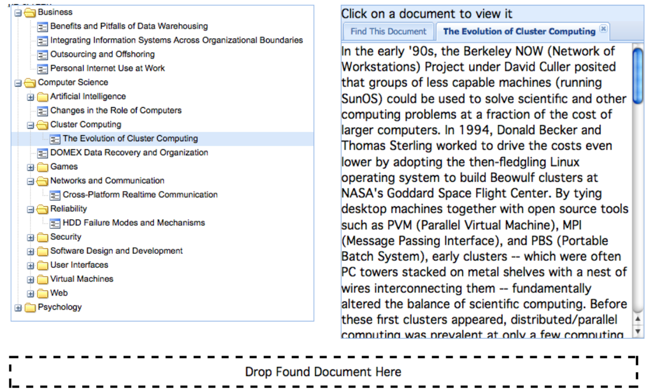

Figure 2: The search tasks interface

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 hierarchy he or she created for an *Imagined Audience*. 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 share less 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* (himself), and the *Consumer* shares common ground with both.

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 for each search task category, corresponding to the number of hierarchies created in each condition of the Organizing phase (see Table 1).

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

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 *audience design* and *common ground* 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 were constructed.

The dependent variable 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 (*M*=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 (20-24 search tasks 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 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 results against the 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. 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 [3].

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
- 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 average.path.length than a hierarchy with 4 or 5 levels and fewer files per folder
- 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 estimating a coefficient for each participant. Using fixed effects allows for greater precision than random effects, but at the expense of statistical power due an increase in the number of degrees of freedom (one for each coefficient in the model). In addition, using random

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

Table 2: Negative Binomial Regression estimates, IRR, % Change. Theta (dispersion parameter) = 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;

effects assumes that effects due to participant-level individual differences are normally distributed [7, p. 650]. This is an important consideration, because while coefficients in a model estimated using fixed effects are not incorrect or biased [7], using random effects can introduce bias in the model predictions by forcing the individual differences to fit a normal distribution when they otherwise would not [16]. 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 [2]. This was one of the stated goals of the analysis, so fixed effects are more appropriate here. Ultimately, this model is a tool for understanding what is going on in this particular dataset; any implications or generalizability based on these results depends more upon the external validity of the experiment than on whether fixed or random effects were used.

The model is constructed as follows:

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

Deciding what regressors to use 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 [7]. I performed two LR tests to compare the goodness of fit between the final model above, and three variations. First, I compared the model above with the "saturated model" which includes one regressor for every observation and 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 above 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).

Model Results

The regression results are detailed in Table 2. Remember that this model uses a log transform of the dependent variable; the model predicts the log of the total.clicks to find a search target rather than the actual count. The coefficients are in the same units and must be transformed back before they can be easily interpreted. For example, consider the estimate for 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 coefficient represents the log of the ratio of the expected number of clicks when 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 coefficient to represent a percentage change in 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%, which means that for each 1-click increase in the shortest path to reach the target file, the total number of clicks to find the search target increases by 20.86%.

A Wald test was performed on each coefficient to test the null hypothesis that the true estimate of the coefficient is zero. Wald tests allow for testing experimental hypotheses, similar in logic to the F-test in ANOVA. These tests are most interesting for the experimentally manipulated factors: imagined.audience, PA.Same, and AC.Same (and the interactions). Table 2 shows that the only significant coefficient 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.

[3]The studentized Breusch-Pagan test was significant (B=91.39, df=56, p=0.0018), indicating heteroskedasticity is present and standard errors are likely underestimated. Table 2 reports White's robust standard errors [7] which are more conservative in the presence of heteroskedasticity; Wald tests on the coefficients reflect the adjusted standard errors, reducing the probability of Type I error.

Model Interpretation

Coefficients 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. Instead of thinking about the Intercept as the value of the dependent variable when all other coefficients are zero, it is actually the $\mathrm{total.clicks}$ for a particular combination of the categorical variables selected to be the baseline by the model. For this model, the baseline assumes $\mathrm{shortest.path}$ and $\mathrm{average.path.length}$ are zero, and the categorical regressors take on the values of *Producer-Audience* (Different), and *Audience-Consumer* (Different). One additional categorical dimension can be layered on top of these: *Imagined Audience* type. The Intercept takes on the value $\mathrm{imagined.audience}$ (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 $\mathrm{imagined.audience}$ (Info. Sci.) in row 3, $\mathrm{AC.Same}$ (Yes) in row 6, and $\mathrm{imagined.audience}$ (Info. Sci.) * $\mathrm{AC.Same}$ (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 $\mathrm{shortest.path}$ (21% increase in $\mathrm{total.clicks}$) and $\mathrm{average.path.length}$ (26% increase in $\mathrm{total.clicks}$).

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* community. Consumers had the the most difficulty when searching in hierarchies created by a *Producer* for a dissimilar *Imagined Audience*. 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 files more easily, regardless of what community they were from, but only when they imagined that they were organizing for somebody like them (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 [6].

Audience Design and Hierarchy Characteristics

The regression model results for the Finding phase of the experiment showed an effect of *audience design*: hierarchies

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		⇒	*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	↓	Intermediate	Same	Different	Different	Info. Sci.
5.	(Intercept)	Worst	↓	Intermediate	Same	Different	Different	Comp. Sci.
6.	−6.01%	Intermediate	=	Intermediate	Different	Same	Different	Comp. Sci.
7.	+2.16%	Worst	↓	Intermediate	Different	Same	Different	Info. Sci.
8.	−4.80%	Intermediate	↑	Worst	Different	Different	Same	Self
9.	−17.76%	Best	↑	Worst	Different	Different	Same	Info. Sci.
10.	−16.75%	Best	↑	Worst	Different	Different	Same	Comp. Sci.

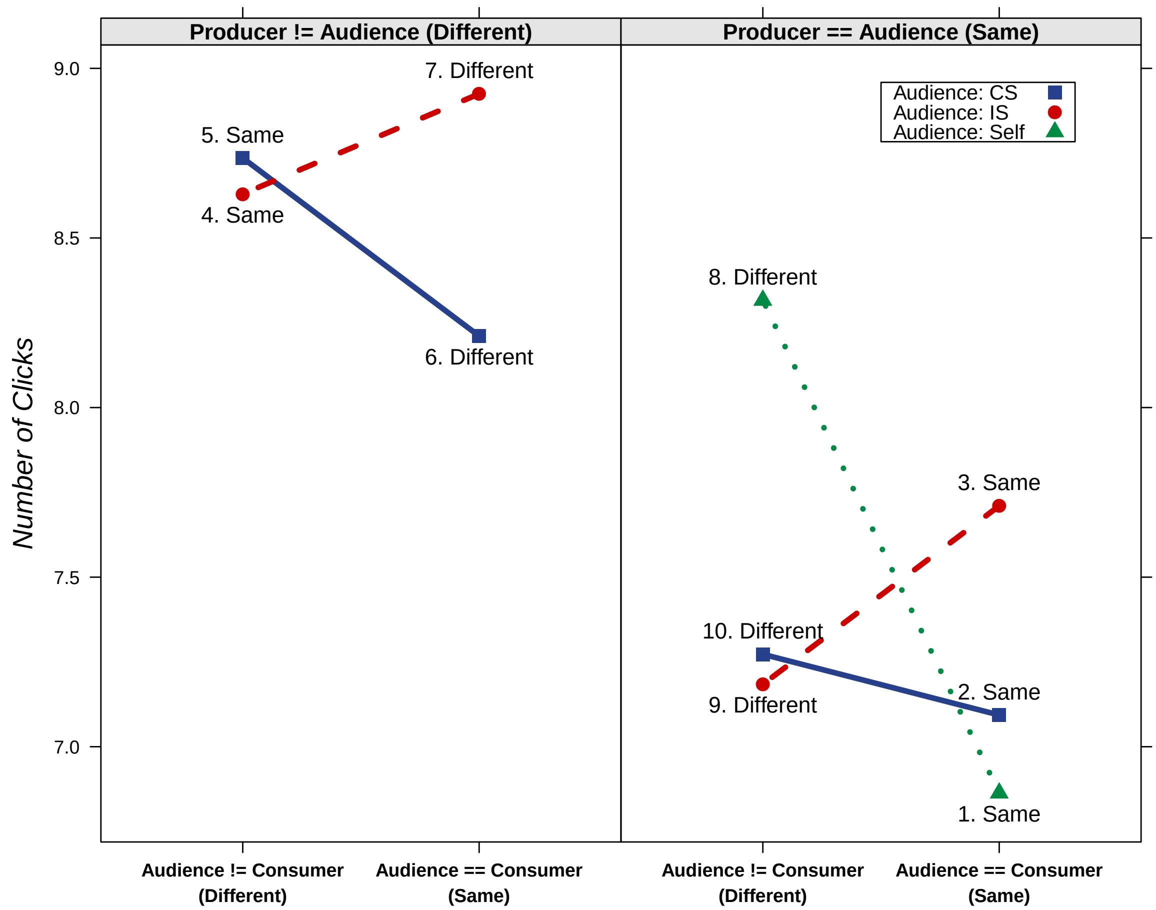

Figure 4: Model results represented as fitted values, based on the median consumer.id **estimate and mean** shortest.path **(2.48 clicks) and mean** average.path.length **(3.93 clicks). "Same" and "Different" point labels refer to *Producer* & *Consumer* community membership, and row numbers in Table 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 [6]. 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.)

Table 4: Statistical comparison of characteristics of hierarchies created for a similar (Same) vs. dissimilar (Different) *Imagined Audience*

	Different (N=374)		Same (N=371)		Kruskal-Wallis Tests	
	M	*SD*	*M*	*SD*	χ^2	
average.path.length	3.92	0.50	3.99	0.65	14.41	***
unique.words	8.24	18.72	3.34	4.33	34.84	***
label.file.similarity	0.26	0.18	0.44	0.20	137.56	***
path.file.similarity	0.10	0.16	0.12	0.15	16.98	***

*** $p < 0.000$

created by participants for an audience similar to themselves were easier for ALL participants to search than hierarchies created for a dissimilar audience. However, it is impossible to tell from the model results alone exactly how audience design affected the hierarchies, causing the pattern of Finding phase results. To better understand the audience design effect, I conducted quantitative analyses on two aspects of the hierarchies, topology and vocabulary:

- *Topology*: measures like average.path.length, reflecting the surface structure of a hierarchy. The word "topology" means both the topographical study of a physical place, or a network of interconnections.
- *Vocabulary*: measures like the number of unique words in a participant's file and folder labels relative to all other participants in the experiment (unique.words), focusing on vocabulary choices participants made when organizing. I also calculated the correlation between file and folder labels and the text of each of the documents the labels represented (label.file.similarity and path.file similarity).

I used these measures as dependent variables in the analyses; the independent variables were pairs of experimentally manipulated variables: PA.Same (Producer-Audience, Same vs. Different), AC.Same (Audience-Consumer, Same vs. Different), and PC.Same (Producer-Consumer, Same vs. Different). I conducted nonparametric Kruskal-Wallis tests because normality and/or homogeneity of variance assumptions were violated in each case. The only independent variable to show significant results for any of the dependent variables was PA.Same; neither the AC.Same nor the PC.Same results were statistically significant.

The results for PA.Same are presented in Table 4. Hierarchies created for an *Imagined Audience* different form the *Producer* contained twice as many unique words in file and folder labels, had file and folder labels that were less correlated with the text of the target files, and were ever so slightly less topographically complex. The issue of statistical vs. practical significance is again relevant for interpreting these results: the difference between the conditions for the average.path.length measure, while statistically significant, is so tiny as to be meaningless. However, the results for the *vocabulary* measures point toward greater word choice variability when hierarchies were created for an *Imagined Audience* different from the *Producer*. It was this variability that led to the Finding phase results. In other words, audience design was helpful only when the *Producer* was familiar with the community for whom he or she created the hierarchy; when the *Imagined Audience* was less familiar, word choices made less sense for finding.

DISCUSSION AND FUTURE WORK

This research was conducted with three goals in mind. The first was to determine whether communication processes are at work in group information management tasks, affecting packaging and finding behavior. Individual users make labeling and organizing choices, and these choices aggregate to form the information structure others use to access files. Therefore, it is important to understand what influences their choices, and how they might be encouraged to make better choices. I found in this experiment that encouraging users to think about the potential audience affects others' finding, and what they are told about the audience makes a difference too. 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 the influence of audience design on hierarchy creation and finding behavior, while replicating previous work. The main finding of this research is that hierarchies created by participants instructed to consider an audience of similar others were easier to search (fewer clicks to reach the search target) than hierarchies created for dissimilar others, regardless of whether the search tasks were completed by participants from the same community as the *Producer* or the *Imagined Audience*. 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 variation. This raises a question, however: why did some of these results differ from predictions based on the literature? Consumers underperformed expectations when Producers tailored their hierarchies for dissimilar others, and did better than expected when Producers organized for similar others. There are two possible reasons for this. First, the audience design manipulation was more complex in this experiment than in [6]; rather than just "self" and "other", this experiment had two different flavors of "other" depending on the intellectual community membership of the participant. Providing additional information to use when doing audience design—both the community membership of the audience and the purpose for the labeling and organizing—could easily have led to different results.

Second, the theoretical predictions assume that audience design improves the communication between two interlocutors. In conversation, common ground accumulates with each utterance; but in an asynchronous "not quite communication" medium like a group information system, there is no support for the accumulation of common ground. One's model of the potential future intended audience is based on a priori beliefs and assumptions, and is not updated through interactions with the system. This means that any discrepancies between one's model of others' knowledge and the other person's actual knowledge are perpetuated without hope of correction.

Because I did not experimentally manipulate or measure participants' mental models of the audience or their knowledge related to the content of the files, I cannot make any specific claims about these discrepancies. However, it is well established that in the absence of information to the contrary, people assume others know the same things they do, and experts overestimate the knowledge of others related to their own area of expertise [11]. Perhaps the instructions of the experiment cause people to try harder than they otherwise would have to imagine what organization scheme might work best for the *Imagined Audience*, but without much information to rely on about that audience, the resulting hierarchies were more difficult to navigate.

The third goal of this research was to gain insight into ways systems might incorporate better support for social aspects of group information management. Given the *audience design* result, and in light of previous research, I suspect that finding ways to incorporate support for the formation of more accurate mental models of other users could help. A bit of practical advice for users of group information systems might be to not get too crazy when organizing and labeling for others; information scent [12] matters, and sticking close to the actual content when labeling and organizing is a better strategy than random speculation about what might work well for someone else. What is most intriguing about these results is that a simple manipulation of experiment instructions brought about the observed differences. Producers imagined someone similar to them while organizing, and this helped *everyone* in the experiment perform better. Likewise, user behavior can be influenced merely by information included in the user interface [14]. These results hint that incorporating information that makes the audience more salient and familiar could help users find the information they need in group information systems. Determining what the right information might be, and how it should be presented to the user, are left for future work.

ACKNOWLEDGMENTS

I thank Rick Wash for writing code for the experiment system; also Libby Hemphill, Morgan Daniels, and anonymous reviewers for comments on an earlier draft, and Judy Olson for feedback on the experiment design. Finally, I thank the University of Michigan Rackham Graduate School for funding this research.

REFERENCES

1. 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.
2. 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.
3. 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.
4. J. M. Carroll. The role of context in creating names. *Discourse Processes*, 3(1):1–24, 1980.
5. 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.
6. 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.
7. D. N. Gujarati. *Basic Econometrics*. McGraw-Hill/Irwin, 4th edition, 2003.
8. 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.
9. 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.
10. 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.
11. R. S. Nickerson. The projective way of knowing: A useful heuristic that sometimes misleads. *CURR DIR PSYCHOL SCI*, 10(5):168–172, 2001.
12. P. Pirolli. Rational analyses of information foraging on the web. *Cognitive Science*, 29(3):343–373, 2005.
13. E. Rader. Yours, mine and (not) ours: Social influences on group information repositories. In *CHI '09*, pages 2095–2098, 2009.
14. 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.
15. 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.
16. G. Verbeke and E. Lesaffre. A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91(433):217–221, 1996.
17. 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.
18. T. Whalen, E. G. Toms, and J. Blustein. Information displays for managing shared files. In *CHiMiT '08*, pages 1–10, New York, NY, USA, 2008. ACM.