

## Table of Contents

1. Introduction .....	2
2. Literature Review .....	5
3. Method .....	12
3.1 Participants and Recruiting .....	12
3.2 Study Procedures.....	13
3.2.1 Search Procedure and Parameters.....	16
3.3 Ethical Concerns .....	17
3.4 Variables and Measurement Techniques.....	18
3.5 Interview.....	20
4. Results .....	22
4.1 Retrieval Task Outcomes .....	22
4.1.1 Participant 1 .....	22
4.1.2 Participant 2 .....	23
4.1.3 Participant 3 .....	25
4.2 Interview Responses.....	26
5. Discussion and Conclusions .....	29
Bibliography .....	36
Appendix A – Mailing List Correspondence .....	38
Appendix B – Post-Task Questionnaire.....	39
Appendix C – Interview Schedule .....	40
Appendix D – Qualtrics Screener Survey.....	41
Appendix E – Informed Consent Form.....	42

## 1. Introduction

In today's workplaces, collaborators often share files using group information repositories which multiple individuals manage and use. These repositories are often large and contain many shared folders for different projects; individuals often participate only in a subset of these shared folders, and within their shared folders they also often interact with only the files and folders relevant to them. Thus, individuals are often more familiar with the files and folders they create and manage or interact with on a regular basis, and less so with files and folders for which this is not the case. When individuals have to interact with a folder structure in a less familiar part of their repository, they may find tasks such as retrieving a file more difficult or frustrating because they are not as familiar with other people's file hierarchies, naming conventions, or project organization as they are with their own. They may thus make mistakes or have to spend more cognitive effort to determine where a file might be or ought to be.

*Group information management* (GIM) is a subdomain of personal information management (PIM) and computer-supported cooperative work (CSCW), and refers to individual actions performed to support collaborative interaction with information items (Lutters et al. 2007). Where PIM refers to management of personal information, GIM focuses specifically on information management in a collaborative setting, and is thus influenced by social and collaborative factors as well as the situation in which information exchange takes place (Lutters et al. 2007). When many users work in a single shared document repository, individual differences in information management strategies (such as filing styles) incur costs of communication, coordination and repository maintenance (Berlin et al. 1993). The range of actions users can take to support

collaborative work (for instance, viewing others' past activity or commenting on a document) is also limited by the technical affordances of the information system used (Dourish et al. 1992). Furthermore, users are more likely to be aware of and interact with the information objects created and placed in the repository by themselves than information objects created and placed there by others (Rader 2009), which may contribute to higher clutter and redundancy in the repository.

*Re-finding* refers to a directed search for information that has been previously encountered (Teevan et al. 2007). Re-finding is often accomplished by locating the object (e.g., file, webpage, email) that contained the information during the previous encounter. Re-finding is distinguished from *finding*, which refers to a directed search for information objects that have not been previously encountered. In the context of personal information management (PIM), re-finding refers to the process of retrieving this information object from within the user's personal space of information (Teevan et al. 2007). A complementary action is *keeping*, or the decision to place or leave an information object in one's personal space of information under the assumption that it will be valuable later. Re-finding is supported by the processes of *recognition*, or the identification of the current context (e.g., folder environment) as the one in which the user has previously identified the retrieval target; and *recall*, or the use of information the user remembers about the retrieval target to assist retrieval (e.g., using remembered metadata to conduct a keyword search).

In collaborative work, an individual may need to go into a pre-existing folder system they are not familiar with in order to retrieve a file they were told about (for example, a new member of an organization who has to fill out a form may be instructed

to look for a template on the organization's network drive). This individual may encounter difficulties finding the file or experience frustration during the process, as their incomplete knowledge of where the file is in the folder hierarchy or how to start looking for it may lead them to spend a lot of time looking in the wrong places or trying to 'work out' where the most likely location of the file might be. The purpose of this study is to investigate the factors influencing how successfully and effectively a user is able to retrieve a file from a shared information repository when they may not be fully aware of its contents or location. The re-finding context was a desktop hierarchical folder structure linked to a cloud-based group information repository (Dropbox).

- *RQ1*: How does familiarity with the retrieval target influence retrieval success and retrieval efficiency?
  - Hypothesis: Familiarity with the retrieval target is positively correlated with retrieval success and retrieval efficiency.
- *RQ2*: What are common characteristics of re-finding experiences in shared folders?

In this study, an elicited personal information retrieval method (detailed in the Literature Review and Section 3.2) is used in conjunction with qualitative methods; as the quantitative data obtained on retrieval outcomes can be linked to the qualitative data gathered from the same participants, this procedure generates explanatory in addition to descriptive information.

## 2. Literature Review

*Recognition and recall.* Prior work by Teevan et al. (2004) and Teevan et al. (2007) suggest two cognitive steps which may assist re-finding. A user employing a *navigation strategy* may become more certain they are going in the right direction if they recognize the context in which the information object was previously encountered. A user employing a *search strategy* may be able to focus their search if they recall specific attributes of the information object they are looking for, such as a keyword in the document metadata (which may or may not be relevant to the specific information experience the user is trying to re-find). This is consistent with the suggestion by Bergman et al. (2014) that users prefer and are more successful at re-finding in their personal spaces of information than in group information repositories due to the greater prevalence of episodic memory cues.

*Preference for navigation over search.* Prior work on re-finding has consistently found that users demonstrate a preference for *navigating* through the information space (following hyperlinks or clicking through the folder hierarchy) instead of *searching* (refining a search engine query or using a desktop search function) (Barreau et al. 1995, Ravasio et al. 2004, Teevan et al. 2004, Bergman et al. 2014). This is also referred to as preference for orienteering over teleporting (Teevan et al (2004)). Explanations for this preference include cognitive ease, a sense of location, and awareness of the context of the retrieval target (Teevan et al. 2004) as well as reliability and avoiding failure or bad performance (Fitchett et al. 2014). Fitchett et al. (2014) further suggest that search is avoided and is a ‘tool of last resort’ due to the cognitive overhead of query formulation and potential information overload from a heavily populated search result list. In the

context of personal information re-finding, this preference may also be motivated by limitations of the search function such as lack of support for full-text search, as well as benefits incurred from being familiar with one's own personal space of information (Ravasio et al. 2004). Applications used to support re-finding strategies include the Recent Documents folder, the Open dialog in specific applications, and Spotlight for Mac OS (Fitchett et al. 2014).

*Wide variance in re-finding behavior between users.* Prior work (Barreau 1995, Teevan et al. 2004, Capra et al. 2004, Fitchett et al. 2014) has demonstrated that information seeking and re-finding behavior varies widely between individuals. Barreau (1995) cites previous work in cognitive science which demonstrates that situational factors significantly influence human cognition; as PIM systems are “more intimate and more in tune with the way people think and work than information systems designed for general use” (Barreau, 1995, p.328), personal context may be one of the significant factors influencing a user's interactions with items in their personal space of information.

Capra et al. (2004) found that in a web search context, users who were familiar with a search task topic or who carried out similar search tasks frequently used search engines less often than those who did not, and suggested this happens because the former group of users have developed more efficient ways of accessing the information than via a search engine. Teevan et al. (2004) discovered a correlation between organization behavior and search behavior, where ‘filers’ who expend considerable effort into organizing their personal space of information are more likely to search or navigate in large steps than ‘pilgers’ who are more likely to navigate in small steps. Fitchett et al. (2014) identify a wide variance in the number of windows opened and the frequency of

window reuse during re-finding. They observed that users who both opened few windows and reused windows infrequently ('minimalists') also retrieved fewer files, and concluded that this approach may not be efficient when file retrieval tasks occur frequently.

*Impact of information context on re-finding outcome.* Attributes of a PIM system's interface, such as what details of a file are shown to a user and how they are displayed, may support recognition and recall and thus influence re-finding. Fitchett et al. (2014) found that users selected their re-finding strategy based on the type of the file to be found, and suggest this is due to differences in retrieval affordances provided by different file types; for instance, multimedia files have little text content and are thus more difficult to search for, and image files may be more quickly located by scanning thumbnails in Quick Look view on the Mac. Bergman et al. (2012) investigated the effects of operating system, interface presentation and folder depth on re-finding success and efficiency. They discovered that folder depth did not significantly affect re-finding and that Mac users retrieve files faster than PC users. They also observed that Mac users tended to have smaller folders and a larger proportion of subfolders compared to PC users, and proposed folder sorting as an additional factor that may influence retrieval outcome. Bergman et al. (2012) also found Icons view leads to faster retrieval than Details view, a finding which was not corroborated by Fitchett et al. (2014) who identified List (analogous to Details view on a PC) and Column view as more efficiently supporting retrieval than Icon and Cover view on the Mac OS. This discrepancy suggests while view layout may influence re-finding outcomes, the specific factors involved are not yet known and may interact, and thus may require further investigation.

As discussed below, group information repositories often provide additional functionality such as change-logging to support collaborative work; the interface and layout of these additional components may influence how much a user knows about a document created or modified by someone else, which may in turn influence the outcome of trying to re-find this document.

*Structures of personal information categorization.* Bergman et al. (2014) outline three major information management strategies: *folder hierarchies*, or a series of nested and mutually exclusive folders; *tags*, which are non-hierarchical and support multiple classification; and *search*, in which users issue queries based on file attributes and which eliminates the categorization step entirely. They further suggest that folder classification is the most preferred structure for organizing personal information, even though creating folders and sorting files into them requires additional effort with little to no improvement in retrieval efficiency. They suggest this preference is independent of search engine quality.

*Technical affordances and restrictions.* As mentioned above, Bergman et al. (2012) demonstrate that interface and OS specifications of a PIM system influence re-finding outcomes. Group information repositories include additional functionality to support collaborative work, such as user permissions, feedback and communication mechanisms, and various forms of change logging (Dourish et al. 1992) or version control (Bergman et al. 2014), which may similarly influence re-finding outcomes.

Dourish et al. (1992) identified two mechanisms provided by group information repositories to support explicit communication and implicit signals of others' actions: *informational*, or mechanisms such as edit logs that allow users to leave notes on their



activity for others, and *role restriction*, which limits the range of actions an individual user is able to perform (for example, write-protection), which reduces uncertainty. They suggest mechanisms for shared feedback (such as a separate simultaneous-edit window designated for communication) which may enable users to observe each other's activity, comment on it, and receive feedback on their own activity; this model of dynamic role negotiation and assignment may reduce both the need for role restriction and the additional effort needed to generate change logs (Dourish et al. 1992).

*Individual differences increase cost of collaboration.* Berlin et al. (1993) suggests "individual information-management strategies do not map well onto group information" (p.1) due to additional costs of communication, coordination and repository maintenance. These costs include 'negotiation overhead', or the need for the group to agree on a filing convention; the additional effort expended by individual users on adhering to agreed-upon conventions, which may differ from their preferred information management strategy; and reduced finding ability over documents they have not encountered. Reduced finding ability can be attributed to: a) not having encountered some of the documents in the repository, which reduces the finder's ability to trigger recall through contextual clues, and b) the possibility of abandoning the search entirely, as the finder does not know for sure the item exists in the repository (Berlin et al. 1993).

Berlin et al. (1993) further discovered that even after categories are agreed upon, users may disagree on the meaning and scope of the categories, which may lead to potential conflict. During their study, they discovered that users of their prototype group information repository eventually circumvented this problem by learning each other's classification rules, which helped them guess where someone else had stored an

information object, although they did not always guess accurately. This workaround is also unintuitive (as remembering and following someone else's ruleset requires additional cognitive work), unfriendly to new individuals, and not scalable to large groups.

*Social dynamics.* Rader (2009) suggests that users interact with information items differently in a shared workspace than they would in their personal spaces of information, in particular that:

1. Users prefer not to modify file or folder structures they did not create.
2. Users are more averse to deleting files.
3. This leads to a more cluttered repository structure, which is more difficult to maintain and search over.
4. Users are often unfamiliar with or unaware of documents outside their own folders, but expect others to be familiar with the items they put into the repository.

Dourish et al. (1992) present the concept of *awareness*, or 'understanding of the activities of others, which provides a context for your own activity', which allows individuals to evaluate how their actions are relevant to and contribute toward group objectives. Dourish et al. (1992) further define the concepts of 'explicitly generated awareness information', which are generally separate from the information object being acted on and which have to be explicitly generated, and 'passively collected and distributed awareness information', which are commonly generated alongside and stored together with the information object. They also found that synchronous CSCW systems (in which users can work collaboratively and view others' activity in real time) were more strongly correlated with passive awareness information while asynchronous CSCW

systems were more strongly correlated with explicitly generated awareness information (e.g. edit logs).

*Methods for investigating re-finding.* Elicited personal information retrieval (EPIR) has been previously used in studies of re-finding in a PIM setting (Bergman et al. 2014). Using this method, the researcher identifies a set of retrieval targets fitting some criteria on the user's computer, generates a list of search primers, and instructs users to retrieve one or more of the identified files while under observation. This method is described by Bergman et al. (2014) as semi-naturalistic: files from an existing naturally occurring group information repository are used, thus increasing the ecological validity of the conclusions drawn from the re-finding task and outcome, while the controlled environment and close observation made possible by an experimental setting is maintained.

Bergman et al. (2014) also describe a more experimental approach in which artificial information items are given to the user and retrieval observed. This has the advantage of greater control over the type and proportion of information objects observed, but may be less ecologically valid as these information objects do not naturally occur in the user's personal space of information and may thus not be typical of an 'authentic' retrieval.

To investigate cognitive processes used, Teevan et al. (2004) and Ravasio et al. (2004) interviewed participants about their experiences with file retrieval. Capra et al. (2005) used a questionnaire to ascertain users' personal contexts and an experiment to assess their retrieval performance. Fitchett et al. (2014) examined logs of user activity

with respect to retrieval time, type of file retrieved, tools used during retrieval, and window activity during retrieval among other logged activity parameters.

*Methods for investigating group information management.* Rader (2009) and Volda (2013) interviewed information workers about their experiences with group information repositories and cloud-based information sharing respectively, while Berlin et al. (1993) developed a prototype group information repository and observed a group of users over the process of adopting the system and adapting to using it. Dourish et al. (1992) enumerated features of three group information repositories and considered whether and how these features fit into their framework of awareness. They also directly observed information workers using a fourth repository and carried out similar analysis.

### **3. Method**

#### **3.1 Participants and Recruiting**

The original goal was to recruit a sample of 15 to 20 participants from the UNC-Chapel Hill student population. The sampling technique used was purposive sampling, with the following criteria required for participation:

- Be a UNC student, at least 18 years old, native English speaker
- Have a computer running Windows 7, 8 or 10
- Be a user of the Dropbox desktop application
- Have at least 50 documents in Dropbox
- Have at least 3 “shared” Dropbox folders that are at least 6 months old in which both you and other people have stored files for collaborative projects

- Be willing to allow the researcher access to your laptop and Dropbox folder during the study

In early March 2016, the researcher sent a recruiting e-mail (Appendix A) to the UNC-CH mass e-mail system and to several departmental mailing lists. Participants were invited to respond to a Qualtrics survey indicating their interest and availability (Appendix D). Upon successful allocation of a timeslot, participants were contacted to inform them of their session time. Participants who completed the session were provided with remuneration of \$20.

By mid-March, only 4 responses had been received, and only 3 respondents met the criteria and were able to be scheduled for participation. Additional recruiting emails were sent to additional departmental mailing lists, but no additional qualified participants were identified. Reflections about the difficulties encountered in recruiting participants for this study are discussed in the Discussion section later in this document.

The results and analyses presented are based on the sample of 3 participants recruited from the UNC-Chapel Hill undergraduate and graduate student population.

### **3.2 Study Procedures**

These study procedures follow the EPIR procedure for investigating re-finding. The study sessions were carried out in the Interactive Information Systems lab and lasted approximately 45 minutes. A typical session proceeded as follows:

1. The participant was greeted and briefed on the session procedures.

2. The participant was invited to sign the informed consent form (Appendix E). The researcher verified that the Dropbox app on the participant's computer matched the requirements of the study (6 months of use and >50 documents).
3. A recording device called the Epiphan Capture Tool was used to make screen recordings of the participant's display. This device has two cables: one that attaches to the video output port (e.g. VGA port) on the participant's computer and another that connects to a USB port on our lab computer. The participant mirrored their display like they would to a display projector, and the Epiphan device sent the video to a recording application on the lab computer. The recordings were stored on a BitLocker-encrypted hard drive partition for security.

The moderator started the recording. The participant was invited to look away while the moderator performed two searches over the participant's Dropbox folder for documents that met the study specifications. The participant was informed that the moderator would only be doing searches to identify files for them to re-find later and no other actions would be performed on their computer. In addition, if the participant wished, that section of the recording would be replayed for them at the end of the session. The search procedure and parameters used are detailed in Section 3.2.1.

4. Based on the two searches, the moderator randomly selected four retrieval targets. Two documents were selected where the document author matches the participant name, and two documents were selected where the document author does not match the participant name.
5. Each search returned a list of documents meeting the search criteria in reverse chronological order. This list was divided into 3 sections, and 1 section was randomly

- selected using a pre-generated random number list. This division and selection process was recursively conducted until a single document was identified. The moderator wrote down the name of each of the identified files on a separate piece of paper. This manual random selection method was chosen over an alternative procedure that would have involved running a Visual Basic Script program on the participant's computer from a USB drive to randomly select the files. The manual search-and-select process used in this study was chosen over the script because it was thought that it would be more transparent and less intrusive to the participant.
6. The participant was handed the paper containing the name of the first retrieval target and was prompted to locate the file starting from the main Dropbox folder.  
Participants were allowed up to five minutes to locate the file. After five minutes, the moderator let the participant know that it was time to move to the next task.
  7. Participants were invited to complete the post-task questionnaire.
  8. Steps 6 and 7 were repeated with the other three retrieval targets previously identified.
  9. The moderator asked a series of questions about the participant's experience finding these files.
  10. The moderator stopped the recording and ask the participant if they would like to watch the section where the moderator conducted searches to identify the retrieval targets.
  11. The moderator thanked the participant for their time, offered them the payment incentive, and asked them to sign a form indicating receipt of the payment incentive.

### 3.2.1 Search Procedure and Parameters

Four searches were done to identify the retrieval targets to be used in the main search tasks.

The moderator invited the participant to start Microsoft Word and navigate to the File -> Info screen, and noted the participant's *document author name*. The moderator then invited the participant to open a File Explorer window and navigate to the Dropbox home folder; this was the starting point from which the searches were conducted.

To identify documents the participant may have authored, the search command

```
(* .docx OR *.xlsx OR *.pptx) System.Author:"[authorname]" AND
datemodified:>1/1/2014
```

was entered using the Windows Search feature; to identify documents the participant may not have authored, the search command

```
(* .docx OR *.xlsx OR *.pptx) System.Author:~!"[authorname]" AND
datemodified:>1/1/2014
```

was used. The search queries comprised three components:

1. *Document format*. The set of retrieval targets was limited to common Microsoft Office document formats.
2. *Author name*. The first search query identifies documents whose authorname matches the document author name previously identified, the second identifies documents whose authorname does not match it.
3. *Date modified*. The set of retrieval targets was restricted to documents modified on or after 1 January 2014.



The moderator alternated authored with non-authored documents. For odd-numbered participants, the first retrieval target was an authored document, and for even-numbered participants, the first retrieval target was a did-not-author document.

### **3.3 Ethical Concerns**

Participants might feel some embarrassment if their Dropbox contains files with sensitive data. To minimize this risk, participants were informed in the recruitment materials and on the informed consent form that they must feel comfortable with allowing their screen to be recorded and allowing the researcher to conduct searches in their Dropbox folder, and that they could change their mind and discontinue participation at any point in the study.

Since the screen recordings made during the session included views of the files and directories on the participant's computer, data security was important. The recordings and data were collected and analyzed on a password-protected computer in the Interactive Information Systems lab in Manning Hall and stored in an encrypted storage format.

The risk of psychological harm was minimal since the search task was designed to resemble searches the participant might ordinarily carry out in daily life. To mitigate this risk, participants were informed in advance about the study procedures, the measures that were taken to protect the data, and the possible risks, and subsequently given the option to discontinue participation in the study. When they indicated their willingness to proceed, they were invited to indicate their acceptance of these risks and willingness to participate in these procedures by signing an informed consent form, and were informed that they could choose to discontinue participation at any time during the study.

### 3.4 Variables and Measurement Techniques

The independent variable in this study is the level of familiarity of the user with the target document to be retrieved. The levels of this variable are defined as follows:

- *Authored*. The user created and worked on the document alone.
- *Co-authored*. The user collaborated with someone else on the document.
- *Different author*. The document was placed in the shared folder by a collaborator.

An initial measurement of this variable was taken for each retrieval target during the retrieval target identification stage—documents were given a preliminary classification of ‘authored’ if they were identified using the search query including the author name, and ‘different author’ if they were identified using the search query excluding the author name. A final measurement of this variable was taken for each retrieval target during the interview stage—documents were classified as ‘authored’ if the participant indicated they wrote all the contents of the file, ‘co-authored’ if the participant indicated they wrote part of the contents and someone else wrote the rest, and ‘different author’ if the participant indicated they wrote none of the contents. In the results, only the final measurement is reported.

The dependent variables measured are the following:

Measures obtained from screen recordings:

- *Retrieval success*. Retrieval of the target document was considered a success.

Retrieval of a different version of the target document was considered a partial success. Any other outcome was considered a failure-to-find.

- *Duration of retrieval.* The time, in seconds, between the first cursor movement and the termination of retrieval.
- *Number of clicks.* The number of mouse clicks made by the participant during a single retrieval session.
- *Number of backtracks.* Use of back button, click on folder in directory path, starting over, using different search terms, opening and then closing a document, switching from search to navigation or vice versa
- *Retrieval strategy.* Whether *navigation* or *search* was employed during retrieval.

Measures obtained from the post-task questionnaire (Appendix B):

- *Perception of difficulty.* How difficult the participant felt the task was to carry out. This was measured using a 5-point scale; the question was “How difficult was it to find this file?”, and the endpoints of the scale were ‘Very Easy’ (1) and ‘Very Difficult’ (5).
- *Perception of confidence.* How confident the participant felt the retrieved document was the target document. This was measured using a 5-point scale; the question was “How confident are you that you found what you were looking for?”, and the endpoints of the scale were ‘Not confident’ (1) and ‘Very confident’ (5).
- *Satisfaction.* How satisfied the participant felt about the search process they used. This was measured using a 5-point scale; the question was “How satisfied were you with the search process you took?”, and the endpoints of the scale were ‘Not satisfied (1) and ‘Very satisfied’ (5).

- *Perception of duration.* Whether the participant felt they took more, less or about equal time to complete the task as they expected to. This was measured using a 3-point scale; the question was ‘Did this take less time, more time, or about the same amount of time you expected?’, and the options were ‘Less time’, ‘About the same time’, and ‘More time’.

Other measurements taken:

- *Folder depth.* The number of levels in the folder hierarchy where the target document was located.

### 3.5 Interview

The question-and-answer stage comprised task-specific questions about the participant’s experience with each retrieval task, as well as more general questions about the participant’s experiences using shared folders. The interview was conducted at the end of the session, after the participant had completed all four tasks. The moderator took notes during these interviews, but the interviews were not recorded.

The task-specific questions were asked about each task and included the following:

- *Consider [filename]. Was searching easy or hard? Why do you think that is the case?* - Participants’ responses to this question provided insight into the folder context in which they carried out retrieval.
- *Did you create this document?* – Participants’ responses to this question eliminates ambiguity caused by the same person using different author names (e.g. across devices).

- *To what extent would you say you wrote the contents of this document?* – Participants' response to this question provides the final measurement for the level of familiarity associated with the document for the given retrieval task.
- *Who put this file in the Dropbox folder?* - Participants' response to this question provides folder context.

The general questions included the following:

- *What do you usually use Dropbox for?* – Participants' response to this question provided insight into folder context and prompted them to elaborate on decisions made about how the shared folder hierarchy was created.
- *What are your typical experiences with shared folders? What was the best experience you have had? The worst? Have you ever had a problem with shared folders? If so, can you describe it?* – This question prompted participants to discuss the benefits and drawbacks they perceived in shared folders generally and Dropbox specifically.
- *When you use a shared folder, how have you managed the files stored in shared folder? Have you ever had to delete files or re-organize files in a shared folder? If so, can you talk about that experience?* – Participants' response to this question provided insight into the situational factors and personal considerations participants thought about when interacting with a shared folder.
- *Describe a time when you worked on a collaborative project and had to file something in someone else's shared folder?* - Participants' response to this question provided insight into the situational factors and personal considerations participants thought about when interacting with a shared folder.

## **4. Results**

In this section, the Results of the study are presented. Due to the small sample size, the results of the three participants will be discussed individually in the sections below. Then, at the end of this Results section, a summary and synthesis of the observations across the participants will be provided.

### **4.1 Retrieval Task Outcomes**

All participants were undergraduate and graduate students at the university, over the age of 18, and native English speakers.

#### **4.1.1 Participant 1**

This participant reported approximately 2 years of Dropbox use, 4 shared Dropbox folders, approximately 1000 documents, approximately 100 folders, and daily use of Dropbox. Table 1 displays the retrieval task outcomes for this participant. The search procedure produced two files the participant authored and two files the participant did not author. The participant successfully retrieved all four files in under 1 minute each. The participant used navigation in three of the four retrieval tasks and search in three of the four retrieval tasks. In all four retrieval tasks, the number of clicks was low (<10), as was the number of backtracks (1-2). The participant found the tasks fairly easy (difficulty rating  $\leq 3$  out of 5), and was generally confident that they found the right file (confidence rating  $\geq 4$  out of 5). They were less satisfied with their search process for the first file (satisfaction rating 2 out of 5) but more satisfied with the process for the other files (satisfaction rating  $\geq 4$  out of 5). They took more time than expected to find

the first and third file, which they also reported as more difficult to find, and the same or less time for the second and fourth file, which they reported as easy to find.

**Table 1.** Summary of results for Participant 1

Time with Dropbox: 2 years; Approx. no. of files: 1000				
Familiarity	Authored	Authored	Did not author	Did not author
Retrieval Performance				
Retrieval success	Success	Success	Success	Success
Retrieval strategy	Navigation -> Search	Navigation	Navigation -> Search	Search
Duration (min)	> 0:34*	> 0:12*	0:43	> 0:06*
No. of clicks	9	4	8	2
No. of backtracks	1	2	2	0
Questionnaire Responses				
Difficulty	3	1	2	1
Confidence	4	5	5	5
Satisfaction	2	5	4	5
Time (vs. expected)	More	Same	More	Less
Interview Responses				
Easy or hard?	Hard	Easy	Moderate	Easy
Explanation	Many files with similar title, unsure of categorization. Nondescriptive name hindered finding	“I thought the content was different from what it was, and went to the wrong folder”	Due to dropbox organization, could have been in multiple folders	Had no idea where it was, so search was the fastest way to find
Who put file there	Participant	Participant	Someone else	Don't know

\*Reported time may be shorter than actual time due to technical difficulties

#### 4.1.2 Participant 2

This participant reported approximately 2 years of Dropbox use, 5 shared Dropbox folders, approximately 4000 documents, approximately 100 folders, and daily use of Dropbox. Table 2 displays the retrieval task outcomes for this participant. The

search procedure produced one file the participant authored and three files the participant coauthored. The participant successfully retrieved all four files in under 20 seconds each. The participant used search in all four retrieval tasks. In all four retrieval tasks, the number of clicks was low (2), as was the number of backtracks (0). The participant found the tasks easy (difficulty rating 1 out of 5) and was generally confident that they found the right file (confidence rating 5 out of 5 for 3 of 4 tasks, and 2 out of 5 for 1 task). They were less satisfied with their search process for the second file (satisfaction rating 2 out of 5) but more satisfied with the process for the other files (satisfaction rating 5 out of 5). They took less time than expected to find all the files, and also reported that the files were easy to find.

**Table 2.** Summary of results for Participant 2

Time with Dropbox: 2 years; Approx. no. of files: 4000				
Familiarity	Authored	Coauthored	Coauthored	Coauthored
Retrieval Performance				
Retrieval success	Success	Success	Success	Success
Retrieval strategy	Search	Search	Search	Search
Duration (min)	00:13	00:16	00:15	00:11
No. of clicks	2	2	2	2
No. of backtracks	0	0	0	0
Folder depth	Unknown (>=2)	Unknown (>=2)	Unknown (>=2)	Unknown (>=2)
Questionnaire Responses				
Difficulty	1	1	1	1
Confidence	5	2	5	5
Satisfaction	5	5	5	5
Time (vs. expected)	Less	Less	Less	Less
Interview Responses				
Easy or hard?	Easy	Easy	Easy	Easy
Explanation	(None given)	(None given)	(None given)	(None given)
Who put file there	Participant	Don't know	Participant	Participant



### 4.1.3 Participant 3

Participant 3 reported approximately 5 years of Dropbox use and approximately 2500 documents. Table 3 displays the retrieval task results for this participant. The search procedure produced two files the participant authored, one file the participant coauthored, and one file the participant did not author. The participant successfully retrieved all four files in under 1 minute for the first three tasks and under 2 minutes for the last task. The participant used navigation in all four tasks and search in one task. The number of clicks was low for the first three tasks ( $<5$ ), as was the number of backtracks ( $= 0$ ) and higher for the last task (clicks = 17, backtracks = 4). The participant found the first and third task easy (difficulty rating 1 out of 5) and the second and fourth task more difficult (difficulty rating 4 out of 5). The participant was generally confident that they found the right file (confidence rating 5 out of 5), but was less confident about the fourth file (confidence rating 2 out of 5). They were generally satisfied with the search process they took (satisfaction rating  $\geq 3$  out of 5). They took less time than expected to find the first two files, but reported that the first file was easy to find and the second hard to find. They took about the amount of time they expected to find the third file, and thought it was easy to find; they took more time than expected to find the fourth file, and thought it was hard to find.

**Table 3.** Summary of results for Participant 3

Time with Dropbox: 5 years; Approx. no. of files: 2500				
Familiarity	Authored	Authored	Coauthored	Did not author
<b>Retrieval Performance</b>				
<b>Retrieval success</b>	Success	Success	Success	Success
<b>Retrieval strategy</b>	Navigation	Navigation->Search	Navigation	Navigation
<b>Duration (min)</b>	0:13	0:42	0:14	1:47
<b>No. of clicks</b>	1	4	3	17
<b>No. of backtracks</b>	0	0	0	4
<b>Folder depth</b>	0	Unknown (>=2)	2	0
<b>Questionnaire Responses</b>				
<b>Difficulty</b>	1	4	1	4
<b>Confidence</b>	5	5	5	2
<b>Satisfaction</b>	5	3	5	4
<b>Time (vs. expected)</b>	Less	Less	Same	More
<b>Interview Responses</b>				
<b>Easy or hard?</b>	Easy	Hard	Easy	Hard
<b>Explanation</b>	Document on top level, could see without navigating File was placed ad hoc; no 'official home'	Ambiguous name; thought the document was about a different subject than it was	Recently accessed and used; knew exactly where it was	Multiple versions of the document; could have been in different folders; common name; hadn't worked on it recently
<b>Who put file there</b>	Participant	Participant	Participant	Participant

## 4.2 Interview Responses

This section will describe participants' responses to the interview schedule (Appendix C), grouped by question topic.

*Primary use of Dropbox.* Participant 1 used Dropbox primarily for work in a research laboratory and stored documents “such as Excel spreadsheets and PowerPoints” in it. Participant 2 used Dropbox for schoolwork and business work, and called it “a space to collaborate with coworkers and classmates”. Participant 3 used Dropbox for

collaborative projects where people need to access the same documents, research, studies conducted with other people, and non-sensitive teaching files.

*Benefits of shared folders.* Participant 1 appreciated the ability to access files from multiple computers and share content for others to view. Participant 2 appreciated the ability to “house files all in one space” and that people could work on the same document in separate locations. Participant 3 also appreciated having “everything in one place”; they also appreciated the ability to access the most updated version of a document, and mentioned this increased trust in the research process and felt more collegial and collaborative.

*Drawbacks of shared folders.* Participant 1 mentioned lack of support for resolving version conflicts— when two people opened a file at the same time and made different changes they would have to go through the file manually and “figure out what from each document to keep”—and mentioned they would appreciate a feature allowed them to see who else was viewing the file at the same time. The Dropbox desktop app has recently introduced a Badge feature to support collaborative editing, and now allows users to see who else is viewing the file, update the local copy of the file to the latest version, and view file history among other functionality. Participant 2 mentioned lack of support for real-time collaborative editing, and made comparisons to other file-sharing systems (Google Docs, OneDrive) which did support this functionality.

*Negative experience.* Participant 1 described an event where, when reorganizing a folder, an accident happened which caused all PC users to lose access to the shared folder. They mentioned that no lasting harm was done and that the group managed to restore appropriate access controls in a timely manner. P2 mentioned having had no

significant negative experiences with Dropbox, and P3 did not mention specific negative experiences.

*Shared file management strategy/folder creation process.* Participant 1's Dropbox contained both personal and shared folders. Participant 2 organized their files into content themes, which made it easy for them and their peers "to know where things belong". Participant 3's shared folder organization was largely group-determined ("depends on the people [they were working with]"); one folder creation procedure was to break up work into sections and create corresponding folders. Participant 3 described the process of group folder creation as natural, unfolding, and made more efficient by a shared research culture.

*How they would reorganize a shared folder.* Participant 3 mentioned they "would not touch other people's stuff" without talking with them first; they would let the file owner know that they would be doing the reorganization and provide a timeframe for the reorganization in case the file owner intended to access the folders or find a file.

*How they would place a new file.* Participant 1 mentioned they would 1) ask the owner of the file in person or by e-mail for permission to place the file; 2) determine where they think the file should go, and 3) place the file. They mentioned that this was made easier because the folders of other people in the lab were often organized similarly and contained similar documents. Participant 3 mentioned that placing a new file was not a usual experience, and mentioned they would be willing to if they could 'tell what it is' or hear from the file owner about it; they would also "make the file name as clear as possible".

## 5. Discussion and Conclusions

This section will discuss the study results and their implications for the two research questions of this study, as well as reflections on the recruiting process and the experimental methodology used in this study. Specific tasks will be referred to using the combination of participant number and task number (e.g. 1.2 for Participant 1's Task 2).

*RQ1: How does familiarity with the retrieval target influence retrieval success and retrieval efficiency?*

Due to recruiting constraints, the sample size was not large enough to conduct quantitative analyses of correlation between level of familiarity and retrieval efficiency. However, qualitative observation of the data suggests that for the three participants studied, the retrieval outcomes may have been influenced more significantly by variations in the shared folder context and participants' own search preferences than by the level of familiarity. Participants 1 and 3 largely started with a navigation strategy (1.1, 1.2, 1.3; 3.1, 3.2, 3.3, 3.4) and occasionally switched to search halfway through (1.1, 1.2; 4.4), while Participant 2 used a search strategy throughout. This may indicate that P2 is a regular user of search rather than navigation, or that they find search to be (in general) more efficient in general. These observations corroborate prior work on re-finding, which suggests re-finding behavior is largely varied and specific to individual users (Barreau 1995, Teevan et al. 2004, Capra et al. 2004, Fitchett et al. 2014), and situationally dependent on both personal information context (Barreau 1995) and the affordances provided by the personal information system (Fitchett et al. 2014).

Berlin et al. (1993) suggest that users are more successful at re-finding in their personal spaces of information than in a group information repository due to differences in information management strategy preferences among individual users, which cannot all be accommodated in a group information repository. Participant 2 mentioned that concerns external to the act of sharing files in a shared folder, such as the amount of storage space available to individuals, influenced group decisions of what files to place or not place in the shared space; they mentioned people might want to limit the number of files placed into a shared folder in order to save space in their own Dropbox. Participant 3 also mentioned knowing users who were uncertain and potentially concerned about whether shared folders occupied space in their personal Dropbox folders.

Participants' responses to Q6 ("Can you describe a time when you... had to put something in someone else's folder?") corroborate the observation of Rader (2009) that people hesitate to interact with others' shared folders, but also indicate that this hesitation does not necessarily translate into an actual unwillingness to interact with an unknown folder system. Specifically, the following conditions mitigated hesitation:

*When collaborators have similar uses for the shared folder, they generate similar file organization strategies and file hierarchies.* Participant 1 mentioned that many individuals in their research lab worked on projects very similar in structure, and thus created similar file hierarchies to organize their files. This may indicate that groups of users who work on similarly structured tasks may develop similar folder hierarchies that correspond to the shared task structure, and thus find each other's folder structure easier to navigate.

*External communication.* When asked about modifying or adding to an existing file structure they did not own, participants 1 and 3 both mentioned contacting the file or folder owner to communicate the intent to make a change, ascertain that the file owner finds the change acceptable, and to better understand where the file should go. Rader (2009) mentions that people do not like to modify shared folders as they are “uncomfortable making assumptions about others’ future information needs” (p. 3).

*Mutually intelligible filing strategy.* Participant 2 mentioned that “each group has a different system” organized into “content themes” in order for all collaborators to understand “where things belong”, and Participant 3 mentioned that shared research culture streamlined the collaborative development of the shared folder structure.

*RQ2: What are common characteristics of re-finding experiences in shared folders?*

All the participants used Dropbox for collaborative work. The participants mentioned access from multiple computers, access by multiple users, and centralized housing for documents as benefits of shared file systems, and that the drawbacks of Dropbox specifically (Participant 1: lack of version control; Participant 2: lack of real-time editing capabilities) do not outweigh the benefits of using it as a file-sharing platform. Dourish et al. (1992) suggest these drawbacks are characteristic of asynchronous file sharing systems and suggest that synchronous file sharing systems (such as Google Docs) may address these concerns through providing users with logs and real-time information about others’ activity in the shared space.

Participants mentioned that the file management strategies employed in shared folders closely mirrored task structure and were determined collaboratively by the group.

This may indicate that groups of users who collaborate on multiple similar projects have shared ideas of task structure and how to develop corresponding folder hierarchies, and thus are able to develop a group folder structure that is easily intelligible and intuitive to everybody. Berlin et al (1993), however, has demonstrated that even when folder organizations are collaboratively developed and agreed upon, differences in filing style and expectations of future use among other factors may still lead individuals to classify items differently. In addition, the cognitive load incurred in navigating an information space organized differently from one's own may reduce the efficiency of re-finding files in these shared spaces.

Participants mentioned that before reorganizing or placing files into other people's shared folders, they would communicate with the folder owner first. Communication between the file owner and the file modifier reduces the assumptive leap the file modifier needs to make about the folder owner's intentions. This may reduce uncertainty over the folder owner's idea of where the file should go, and may also reduce a user's unwillingness to interact with the folder since permission has been obtained. However, the necessity of additional external communication also incurs additional cost in terms of the additional effort users must make, and the communication infrastructure which must first be established. Dourish et al. (1992) present some ways in which groups may carry out this communication (referred to as *explicitly generated awareness information*), such as additional shared windows in the shared folder for users to read and write in. The Dropbox desktop app supports the creation of this awareness information through its Badge feature, which allows users to comment on others' files and notifies the file owner of comments.



### *Reflections about recruiting*

Several factors may have contributed to the recruiting issues faced. Firstly, as the inclusion criteria were stringent, the pool of potential participants may have been more limited than anticipated. Furthermore, people may have been deterred from participating out of unwillingness to allow their Dropbox to be viewed or their desktop to be recorded. In addition, the study was conducted during a busy time of the academic year.

### *Elicited Personal Information Retrieval Study Method*

The study was carried out as an experiment using the EPIR procedure for investigating re-finding in a PIM setting. Using this method, the researcher identified a set of retrieval targets fitting specific criteria on the user's computer, generated a list of search primers, and instructed users to retrieve one or more of the identified files while under observation.

*Advantages.* As EPIR is conducted over a naturally occurring group information repository, the results of the experiment are likely to be fairly generalizable and ecologically valid. At the same time, the researcher is able to exert control over the experimental conditions and thus increase reliability of results.

*Disadvantages.* During the preliminary file search, files were separated into 'authored' and 'did not author' categories. This strategy does not always fully capture the extent to which the participant actually worked on the file: for instance, it identified files in the correct category for Participant 1, and was less effective at doing so for Participant 2. Thus, using this method of preliminary file categorization, the researcher cannot control for the number of coauthored files relative to the number of 'different author'

files. Repeated searching over the same set of folders may also lead to learning effects. For instance, a participant may be able to locate a file not because they ordinarily remember its location, but because they encountered it during a previous search task carried out during the study.

The study captures only a snapshot of the participants' retrieval behavior. A longitudinal study using naturalistic observation may provide richer and more detailed insight into participants' document retrieval patterns and habits.

Due to study constraints, the specific *level of involvement* the participant had in a 'coauthored' document was not considered in this study. Potential future work may investigate whether a participant who worked extensively on a document is more likely to find it than a participant who has not worked extensively with it.

More generally, the participant sample may show self-selection bias because participants must consent to their files being accessed in order to participate in the study. Participants may also carry out retrieval differently under observation and in a controlled experimental setting than they would under ordinary conditions. For instance, in the study, they were unable to contact collaborators to ask after the location of the file.

*Preliminary identification of files based on authorname search.* This method of identifying retrieval targets was largely successful in the case of Participant 1 (identified 2 documents authored by the participant and 2 documents not authored by the participant) and less successful in the case of Participant 2 (identified 1 document authored by the participant and 3 co-authored document). Possible drawbacks of using authorname presence as a preliminary proxy measure for familiarity include that users may have different author names on different computers, and that users who work closely with

others may have very few documents that were wholly authored by themselves or by someone else.

### *Conclusions*

Re-finding in a GIM setting depends on a number of complex and interacting factors. From the data, the three participants mentioned a range of individual and contextual factors and considerations that influence how they act in their shared folders, including task structure and collaboration history with the group. In at least one instance, complete unfamiliarity with the research target and no knowledge of its location led to a choice of a more efficient strategy (searching). Possible avenues for future research may include: Do individuals choose their search strategies based on which shared folder they know or believe the file belongs to?

## Bibliography

- Barreau, D. K. (1995). Context as a factor in personal information management systems. *Journal of the American Society for Information Science*, 46 (5), 327-339.
- Barreau, D., & Nardi, B. A. (1995). Finding and reminding: file organization from the desktop. *SIGCHI Bulletin*, 27 (3), 39-43.
- Bergman, O., Whittaker, S., Sanderson, M., Nachmias, R., & Ramamoorthy, A. (2012). How Do We Find Personal Files?: The Effect of OS, Presentation & Depth on File Navigation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. May 5-10, 2012, Austin, Texas, USA. New York, NY: ACM.
- Bergman, O., Whittaker, S., & Falk, N. (2014). Shared files: The retrieval perspective. *Journal of the American Society for Information Science and Technology*, 65 (10), 1949-1963.
- Berlin, L. M., Jeffries, R., O'Day, V. L., Paepcke, A., & Wharton, C. (1993). Where did you put it? Issues in the design and use of a group memory. *Proceedings of the INTERACT '93 and CHI'93 Conference on Human Factors in Computing Systems*. April 24-29, 1993, Amsterdam, The Netherlands. New York, NY: ACM.
- Capra, R., & Pérez-Quiñones, M. A. (2005). Using Web Search Engines to Find and Refind Information. *IEEE Computer* 38 (10), 36-42.
- Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared workspaces. *Proceedings of the 1992 ACM Conference on computer-supported cooperative work*. November 1-4, Toronto, Ontario, Canada. New York, NY: ACM.
- Fitchett, S., & Cockburn, A. (2014). An empirical characterisation of file retrieval. *International Journal of Human-Computer Studies*, 74, 1-13.
- Lutters, W.G., Ackerman, M.S., and Zhou, X. (2007). Group Information Management. In Jaime Teevan and William Jones (Eds.), *Personal Information Management* (22-34). Seattle: U of Washington.
- Jones, W. and Teevan, J. (2007). Introduction. In Jaime Teevan and William Jones (Eds.), *Personal Information Management* (22-34). Seattle: U of Washington.
- Rader, E. (2009). Yours, Mine, and (Not) Ours: Social Influences on Group Information Repositories. *Proceedings of the ACM Conference on Human Factors in Computing Systems*. April 4-9, 2009, Boston, MA, USA. New York, NY: ACM.

- Ravasio, P., Schär, S.G., & Krueger, H. (2004). In pursuit of desktop evolution: User problems and practices with modern desktop systems. *ACM Transactions on Computer-Human Interaction*, 11(2), 156-180.
- Teevan, J., Alvarado, C., Ackerman, M. S., & Karger, D. R. (2004). The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. *Proceedings of the ACM Conference on Human Factors in Computing Systems*. April 24-29, 2004, Vienna, Austria. New York, NY: ACM.
- Teevan, J., Capra, R., & Pérez Quiñones, M. (2007). How People Find Personal Information. In Jaime Teevan and William Jones (Eds.), *Personal Information Management* (22-34). Seattle: U of Washington.
- Voida, A., Olson, J. S., & Olson, G. M. (2013). Turbulence in the Cloud: Challenges of cloud-based information work. *Proceedings of the Conference on Human Factors in Computing Systems*. April 27-May 2, 2013, Paris, France. New York, NY: ACM.

## Appendix A – Mailing List Correspondence

Subject: Study about Dropbox

I am conducting a study about how people find files stored on Dropbox and am looking for UNC students to participate in the study. Participation will take approximately one hour and you will be offered \$20 for your time. Sessions will be held in Manning Hall (next to Lenoir).

To qualify as a participant, you must:

- Be a UNC student
- Be at least 18 years old
- Be a native English speaker
- Have a computer running Windows 7, 8 or 10
- Be a user of the Dropbox desktop application (this means that you have a Dropbox folder on your computer, not just used it through the web)
- Have at least 3 “shared” Dropbox folders that are at least 6 months old in which both you and other people have stored files for collaborative projects
- Have at least 50 documents in Dropbox
- Be willing to allow the researcher access to your laptop and Dropbox folder during the study session (no changes will intentionally be made to your files or folders).

You will be asked to allow a screen capture device to be plugged in to your computer and we will record your screen during the study session. Next, we will ask you to look away for a few minutes while we select several files in your Dropbox folder that we will ask you to try to re-find (no changes will intentionally be made to your files or folders). We will then ask you to try to find these files while we record your actions. If you are not comfortable with the researcher interacting with the contents of your computer’s hard drive/Dropbox and having recordings made of this, you should NOT participate in this study.

If you are interested in participating, please fill in this Qualtrics survey ([https://unc.az1.qualtrics.com/jfe/form/SV\\_2bnkc4pbdudruJf](https://unc.az1.qualtrics.com/jfe/form/SV_2bnkc4pbdudruJf)) indicating your availability and the researcher will contact you shortly.

Thank you, and have a nice day!



## Appendix C – Interview Schedule

[For each retrieval target]

- Consider [filename]. Was searching easy or hard? Why do you think that is the case?
- Did you create this document?
- To what extent would you say you wrote the contents of this document?
- Who put this file in the Dropbox folder?

[General questions]

- What do you usually use Dropbox for?
- What are your typical experiences with shared folders? What was the best experience you have had? The worst? Have you ever had a problem with shared folders? If so, can you describe it?
- When you use a shared folder, how have you managed the files stored in shared folder? Have you ever had to delete files or re-organize files in a shared folder? If so, can you talk about that experience?
- Describe a time when you worked on a collaborative project and had to file something in someone else's shared folder?



# Appendix D – Qualtrics Screener Survey

Name

E-mail address

Are you a UNC student over the age of 18?

- Yes
- No

Are you a native English speaker?

- Yes
- No

During which of the following times will you be available to attend a study session?

- (Example only) Mon, Jan 11, 9:00am-10:00am
- (Example only) Tues, Jan 12, 9:00am-10:00am
- (Example only) Wed, Jan 13, 9:00am-10:00am

How many shared Dropbox folders do you have in which both you and other people have stored files for collaborative projects?

Roughly how many folders do you have in Dropbox?

How frequently do you use Dropbox?

- Once a year or less
- A few times a year or less
- A few times a month
- A few times a week
- Daily

Have you installed the Dropbox application on your computer? This means that you will have a Dropbox folder on your local machine.

- Yes
- No

How long have you had Dropbox installed on your computer?

What is the primary activity for which you use Dropbox collaboratively?

## **Appendix E – Informed Consent Form**

### **University of North Carolina at Chapel Hill Consent to Participate in a Research Study Adult Participants**

**Consent Form Version Date:** 12/17/15

**IRB Study #** 15-3269

**Title of Study:** Investigating the effect of familiarity with target document on retrieval success in group information repositories

**Principal Investigator:** Kimberly Hii

**Principal Investigator Department:** School of Information and Library Science

**Principal Investigator Phone number:** 919-360-8184

**Principal Investigator Email Address:** kimhii@live.unc.edu

**Faculty Advisor:** Robert Capra

**Faculty Advisor Contact Information:** 919-962-9978

---

#### **What are some general things you should know about research studies?**

You are being asked to take part in a research study. To join the study is voluntary.

You may refuse to join, or you may withdraw your consent to be in the study, for any reason, without penalty.

Research studies are designed to obtain new knowledge. This new information may help people in the future. You may not receive any direct benefit from being in the research study. There also may be risks to being in research studies.

Details about this study are discussed below. It is important that you understand this information so that you can make an informed choice about being in this research study.

You will be given a copy of this consent form. You should ask the researchers named above any questions you have about this study at any time.

#### **What is the purpose of this study?**

The purpose of this research study is to investigate how people re-find files on shared Dropbox folders. You are being asked to be in the study because you have indicated that you have used the Dropbox desktop application for collaborative purposes for over 6 months and that your Dropbox folder contains at least 50 documents.

#### **Are there any reasons you should not be in this study?**

You should not be in this study if you have not used the Dropbox desktop application for collaborative purposes for at least 6 months, your Dropbox folder contains less than 50 documents, you do not want your computer screen to be recorded, or you do not wish to allow the researcher temporary access to your laptop or Dropbox folder.

**How many people will take part in this study?**

There will be approximately 25 people in this research study.

**How long will your part in this study last?**

Your part in this study will last approximately 1 hour. No follow-ups will be necessary.

**What will happen if you take part in the study?**

1. You will be briefed and invited to sign this informed consent form.
2. The researcher will verify that the Dropbox app on your computer matches the requirements of the study.
3. The researcher will connect your computer to a recording tool using the VGA, HDMI, or display port on your computer. This tool will record everything that is shown on your computer screen during the session.
4. The researcher will conduct several searches in your Dropbox folder for documents which meet the study specifications. For the purposes of this study, you will not be allowed to view this search, but this process will be recorded and you will be given the option to review it at the end of the session. When the researcher conducts the searches, no changes will be made to your files or folders. The purpose of the search is only to identify files to ask you to re-find in the next step of the study.
5. Next, the researcher will ask you to find about four files in your Dropbox by showing you their filenames. After finding each one, you will be asked to complete a post-task questionnaire.
6. After finding the files, the researcher will ask you a series of questions about your experience finding these files, and about your experiences using Dropbox for collaborative projects.
7. The researcher will give you an opportunity to review the recording.

By participating in this study, you consent for your computer screen to be recorded for the duration of the study session.

**What are the possible benefits from being in this study?**

Research is designed to benefit society by gaining new knowledge. You will not benefit personally from being in this research study.

**What are the possible risks or discomforts involved from being in this study?**

Participants may feel some embarrassment if their Dropbox contains files with sensitive data. Recordings will be made of your computer screen that will include information about the files stored on your computer. There may be uncommon or previously unknown risks. You should report any problems to the researcher.

**How will information about you be protected?**

The screen recordings made during the session will be stored on a password-protected computer using an encrypted storage format. In all data analysis and subsequent reports, participants will be identified only by a randomly assigned participant ID.

Participants will not be identified in any report or publication about this study. Although every effort will be made to keep research records private, there may be times when federal or state law requires the disclosure of such records, including personal information. This is very unlikely, but if disclosure is ever required, UNC-Chapel Hill will take steps allowable by law to protect the privacy of personal information. In some cases, your information in this research study could be reviewed by representatives of the University, research sponsors, or government agencies (for example, the FDA) for purposes such as quality control or safety.

**What if you want to stop before your part in the study is complete?**

You can withdraw from this study at any time, without penalty. The investigators also have the right to stop your participation at any time. This could be because you have had an unexpected reaction, or have failed to follow instructions, or because the entire study has been stopped.

**Will you receive anything for being in this study?**

You will be offered \$20 upon completion of your study session.

**Will it cost you anything to be in this study?**

It will not cost you anything to be in this study.

**What if you are a UNC student?**

You may choose not to be in the study or to stop being in the study before it is over at any time. This will not affect your class standing or grades at UNC-Chapel Hill. You will not be offered or receive any special consideration if you take part in this research.

**What if you have questions about this study?**

You have the right to ask, and have answered, any questions you may have about this research. If you have questions about the study (including payments), complaints, concerns, or if a research-related injury occurs, you should contact the researchers listed on the first page of this form.

**What if you have questions about your rights as a research participant?**

All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject, or if you would like to obtain information or offer input, you may contact the Institutional Review Board at 919-966-3113 or by email to IRB\_subjects@unc.edu.

**Participant's Agreement:**

I have read the information provided above. I meet all the eligibility criteria. I agree to allow the moderator to have access to my computer and my Dropbox folder for the purposes of the study outlined above. I consent to have my computer screen recorded during the study. I have asked all the questions I have at this time. I voluntarily agree to participate in this research study.

---

Signature of Research Participant

---

Date

---

Printed Name of Research Participant

---

Signature of Research Team Member Obtaining Consent

---

Date

---

Printed Name of Research Team Member Obtaining Consent