That syncing feeling: Early user experiences with the cloud

Cathy Marshall **John C. Tang**
Microsoft Research
1065 La Avenida
Mountain View, CA 94043 USA
cathymar | johntang@microsoft.com

ABSTRACT

We studied how people use file sync and sharing services to better understand how early adopters conceptualize their interactions with the cloud. A survey of 106 users provided background information about current use of these cloud storage services and identified 19 people for in-depth interviews. Use cases described in the interviews revealed a hierarchy of concepts that participants needed to master to make full use of these services. Five pivotal concepts demonstrate that users make sense of the cloud as a: personal file repository, shared file repository, personal replicated file store, shared replicated file store, and synchronization mechanism that coordinates among replicas. We propose specific ways in which process transparency and interface scaffolding can help users build a more robust model of cloud services.

Author Keywords
File sync, file sharing, cloud user experience.

ACM Classification Keywords
H5.3. [Information Systems]: Information interfaces and presentation---Group and Organization Interfaces

General Terms
Human Factors.

THE USER EXPERIENCE OF THE CLOUD

While the popularity and potential of cloud computing continues to grow, much of the work to date has focused on the technical infrastructure, such as optimizing throughput, scaling up in capacity, and maximizing uptime [2]. As cloud computing continues to deliver more services to the end user, we wanted to understand how users make sense of their interactions with the cloud. In many cloud-based applications, like email, the users are focused more on the application than on interacting with the cloud that provisions it [9]. Thus, they reap the benefits of accessing the data and services they want from any device without having to learn how to interact with the cloud.

DIS 2012, June 11-15, 2012, Newcastle, UK.

Meanwhile, cloud-based storage is becoming the backbone of diverse services for sharing and synchronizing files. Services such as Dropbox and Google Documents enable people to use the cloud to edit and share files across multiple computing devices or with other people via a network of servers. Because a cloud infrastructure is essential to these services, it exercises early conceptions of what the cloud is and what it can do for people. We expect that accurate and robust conceptual models [10] will be important in helping users make the most of cloud services. By exploring how people are using these early cloud-based end user services, we hope to gain insight that will help guide the development of future interfaces to the cloud.

We sought to explore three basic questions about early use of these cloud storage services. First, how are early adopters using the services to sync and share their files? As people describe how they (and their collaborators) use the services, they reveal their understanding of where content is stored and how it gets there. Second, what are the conceptualizations (and misconceptions) that accelerate or inhibit peoples' use of these services? Finally, how can interaction design be used to shape users' understanding of the cloud and allow them to develop a more accurate picture of what the services can do for them?

PRIOR RESEARCH ON FILE SYNC AND SHARING

A growing body of research has demonstrated users' need to synchronize files across devices. Dearman and Pierce [7] showed that many people routinely use several computing devices, and often need to have access to information and files that are on one of their other computing devices. While they found some use of early file synchronization services, they also observed that users had difficulty trusting them.

Sohn et al., [15] focused on how the increased use of mobile devices added to the challenge of getting access to the information needed from those devices. Karlson et al. [11] observed that tasks often flow across multiple devices (including from a mobile device to a computer), further emphasizing the need for cross-device access. These studies identified the need for file sync services that offer access to any file from any device.

Studies of file sharing systems laid the groundwork for these results, including the BSCW system [4] which used the Web to enable people to easily share files over the

internet. But Rader's study of group information repositories [14] showed that people manage files in a shared repository differently than in their own file directories, indicating some conflicting conceptual models of how to share files. People were reluctant to delete files that they did not own, but disliked the resulting clutter which diminished the shared repository's value for everyone. These studies raise some of the interface complexities around file sharing.

These user needs identified by prior research are being met by recently developed cloud-based file sync and sharing services. While Dropbox and Google Docs are two of the most popular, they compete with a wide range of services with different features and cost structures. But, as Bowers points out [5], it is not just the technology but the work practices that develop around it that shape what users can actually accomplish. We set out to study the early use of these services to understand the models users were developing as they interacted with the cloud.

STUDYING FILE SYNCING AND SHARING

We conducted a two-phase study of cloud-mediated file syncing and sharing. First we piloted and administered a 106-person survey to gather background data and identify prospective participants for the second part of the study, which consisted of 19 in-depth interviews. The survey included 15 questions about respondents' use of syncing and sharing software and services and six demographic questions, and took about fifteen minutes to complete. We recruited respondents for voluntary participation via posts in Facebook, Twitter, and email groups that spread beyond our personal contacts. Respondents ranged from computer research professionals to moms using shared files to coordinate school volunteer tasks.

For the interviews, we selected 19 of the 106 survey respondents for a balance of characteristics such as age, gender, software used, level of technical sophistication, satisfaction with their syncing and sharing solutions, and apparent extent of the demands they placed on the cloud. Semi-structured, open-ended interviews were conducted over Skype to reach participants in their usual computing environments so they could access their file syncing and sharing services. To ground the interviews, we asked the participants to view their synced folder or cloud document repository. Interviews lasted about an hour and were audio-recorded for later review and transcription. Participants received a $25 gift card for their time.

Our interview data were first open-coded for recurring themes, then analyzed by comparing specific instances of those themes across participants [16]. Our analysis focused on how they used syncing and sharing services, and on the conceptual models they were forming through their usage.

Table 1 summarizes the characteristics of the interview participants from their survey responses. Participant IDs assigned based on the survey responses are used to identify quotations and stories from their interviews.

Survey Results

The 106 completed surveys provided background information about respondents' current use of syncing and sharing services. Of the respondents, 60% were male, 37% were female. Many were in the 20-35 yr. range (39%), but several were under 20 (3%) and over 65 (2%). Each

ID	Gender	Age group	Technical level	# of devices	Syncing Satisfaction	Extent of sharing	Sharing Satisfaction
S006	F	51-65	low	2	4	2-5 groups	6
S007	M	36-50	medium	2	2	>5 groups	2
S012	F	51-65	medium	3	2	2-5 groups	6
S014	M	20-35	high	3	6	2-5 groups	3
S017	M	20-35	high	6	6	2-5 groups	5
S018	F	36-50	low	6	3	2-5 groups	6
S032	M	36-50	medium	5	7	2-5 groups	7
S038	M	20-35	medium	4	5	2-5 groups	4
S039	F	51-65	medium	6	7	none	7
S051	F	36-50	low	3	4	>5 groups	4
S052	M	51-65	high	6	6	2-5 groups	5
S055	M	36-50	medium	4	6	2-5 groups	5
S058	F	20-35	low	8	7	2-5 groups	7
S062	F	20-35	low	3	4	2-5 groups	5
S065	M	20-35	high	4	2	2-5 groups	4
S078	M	36-50	high	3	7	2-5 groups	7
S092	M	20-35	high	4	5	>5 groups	5
S102	F	51-65	medium	4	2	1 main group	5
S106	M	36-50	high	6	6	2-5 groups	5

Table 1. Interview participant characteristics.

respondent used an average of 4.4 computing devices, including workplace and personal desktops, laptops, slates, and smartphones.

Most respondents used cloud-based services to both sync files across devices and share files with other people. Only 6% of the respondents said that they did not use a service for file sharing, while 19% said that they did not sync files using a service. Regarding how much they used each service, 57% said that they used Dropbox frequently, followed by Google Docs at 47%. Most of the respondents (63%) rated two services as being used "frequently" or "occasionally", with only 22% saying that they only used one service at those levels and the rest used three or more services. In addition to the most popular services, respondents mentioned Microsoft LiveMesh, Apple MobileMe/iCloud, SugarSync, Evernote, SyncToy, Microsoft OneNote, FilesAnywhere, and ShareFile. Many users were familiar with multiple syncing and sharing services, which gave them an opportunity to compare how they used different services and explain why they chose a particular service for specific activities.

The survey asked how satisfied respondents were (on a 7-point scale where 7=very satisfied) with their file syncing solution as well as their file sharing solution. The average response for both questions was 5.1, although the median response was 6 for file syncing, compared to 5 for file sharing. While these satisfaction ratings are positive, they also show that there is room for improvement.

Examplars: Dropbox, Google Docs, and iCloud

Although respondents mentioned various cloud services, we focus on three, Dropbox, Google Docs, and iCloud, as exemplars. Describing the differences among them will serve as groundwork for the rest of the paper.

Dropbox (https://www.dropbox.com/) is a service that synchronizes file folders among all devices registered with a common account, using the cloud as an intermediary. A Dropbox folder looks like an ordinary file folder, but it is more accurate to think of it as a local cache of a repository stored in the Dropbox cloud. Dropbox subfolders may be shared with designated users, so that they can synchronize a local folder with the files most recently stored in the cloud. Any file type can be stored, synchronized, and shared via Dropbox. If work is done on a Dropbox file while the device is offline, those updates are synchronized when the device comes back online. If concurrent work is done on a file from different devices, the resulting multiple copies must be manually reconciled.

Google Docs (https://docs.google.com/) works as a cloud repository that can be accessed by any web-connected device. Users must have network access to edit the current version of the file. Users can share files with others, causing them to appear in their list of Google Docs, and even concurrently edit content (documents, spreadsheets, presentations, etc.) within Google applications. Other file types may also be stored and shared through the service.

Apple's iCloud (http://www.apple.com/icloud/) is a service that seamlessly and wirelessly synchronizes information used by Apple applications (e.g., calendar, contacts, photos, iWork) across Apple devices. Since iCloud integration happens at the application level, the service controls how much of the data synchronization is presented to the user. Currently, iCloud does not offer file sharing with others.

USES OF SYNC AND SHARING SERVICES

During the interviews, users described different ways they used cloud-based file sync and sharing services in both their professional and personal lives. The survey had led us to believe that respondents used these services to share files slightly more often than they used them to sync files, and that they were reasonably satisfied with these services. But the interviews revealed that this was only part of the story.

The interviews elicited variations of our two basic use cases with some subtle but important distinctions. Our broadened picture of use cases included transferring files between devices or people; backing up active work; sharing files in a cloud repository; syncing files among personal devices; and syncing files shared with others. We explore each use in turn, describing how it contributes to users' development of a conceptual model of the cloud.

Transfer Between Devices

Although services like Dropbox, FilesAnywhere, and SugarSync can actively sync personal folders among devices, some participants used these services to merely transfer files from one device to another like they would use a USB stick. Once the files reached their destination, they were removed from the folder, thus eliminating the possibility of syncing future changes to the file. For example, S102 used Dropbox to move files between personal devices: "*If … I want to get some stuff from my desktop to my laptop up in the living room, that's the fastest way to do it now. I could put it on a [USB] stick obviously*."

Transfer Among People

Because these services accommodate large files, some participants used them as a substitute for email attachments. Collaborators did not expect to accumulate a shared repository; rather, one person would put the file on the cloud store and the other(s) would claim it (i.e., the folder literally serves as a dropbox). For example, S078 used Dropbox to share movies with his girlfriend:

> *When she goes into Dropbox and there's a new movie there, and she knows to take it out so that there's space for new movies to come in. We use it as a transfer portal... I don't think she has a deep understanding of the server in the backend is doing all that work.*"

Transferring files to other devices or people through the cloud filled a practical user need, but did not help users learn to use any other cloud-specific features or services.

Active File Backup

Backing up active files to the cloud is seen as transferring files to the future (when one has inadvertently destroyed the local copy) or to a potentially unknown destination, such as a computer not yet identified. Some participants selectively backed up specific files, while others put all of their active files in the cloud. For example, S007 explained that one of his personal Dropbox folders contained "*stuff that I didn't want to lose. Some of it was things when I was traveling. And some of it is presentations which I just put in Dropbox in case I was unable to use [my own laptop]*."

Cloud-based File Sharing

Unlike file transfer, which suggests that the files stored in the cloud are transient and not synced between devices, file sharing relies on a persistent cloud-based repository. Participants liked having a place to share files with other people that could be accessed from anywhere; they could rely on the cloud store to have the most recent version and to support longer-term projects that involved sustained collaboration. For example, S018 explained how she shared files with others teaching with her: "*I open up Dropbox, and up comes my set of files… [for] that particular class. I would open up whichever day is relevant.*"

Sometimes a single cloud-based file is used to accumulate bits of information across multiple people to coordinate the group's activities. The concurrent editing feature in Google Docs supports this kind of file sharing without the possibility of introducing conflicts, even though users rarely engage in the kind of synchronous session-based collaboration anticipated by shared editing tools [12]. Almost all of the study participants had used Google Docs for well-structured and timely tasks such as filling in a cell or a column in a spreadsheet. S012 reported a typical activity of this sort, "*organizing the information for a youth soccer team in a big Google Docs spreadsheet. Every family just inputs their address information, and their kids' uniform sizes, and all that kind of stuff.*"

Syncing Files Among Personal Devices

To many participants, syncing files among their devices was not an intuitive concept when they began using Dropbox. After getting introduced to Dropbox via one of the previously described uses, it often took experimentation over time to realize that the folder in Dropbox not only offered access to files in the cloud, but also enabled editing and updating across different computers. S038 explained:

> *It took me a while to learn that I had to leave something in a folder and not mess with it, or use a folder that would sync through Dropbox as a place where a document lives in order for me to consistently use it across machines.*

We asked participants who had discovered the benefits of synced folders whether they would like to store their entire MyDocuments folder on Dropbox. While this decision is largely governed by the practical aspects of Dropbox's free storage quota, it also raised privacy and security issues that we discuss in a later section.

Using Synced Folders Collaboratively

Maintaining a collaborative synced store was the most conceptually complex use of cloud-based services. They had to agree with their collaborators to share folders this way, negotiate a workable shared structure (since the folder hierarchy is the same for all), and coordinate interactions so that changes propagated to everyone's local computers predictably. These concepts are tricky to understand and enact. If collaborators had different models of how synced shared folders worked, surprising, even conflicting, behaviors often ensued. Beyond the negotiation of shared structure described in prior work, e.g. [3], we focus on the propagation of changes and conflict resolution.

For a family project, S012 used Dropbox to create a shared folder where she and her brother could both upload photos. As we interviewed her, she realized there were both situations where she did and did not want to allow users to delete files from others' computers, and that even a notion of file ownership would not resolve that tension:

> *I put stuff in it. It goes into the cloud. And then it seems like it, in the background, copies that down to the other peoples' file systems in a folder called 'My Dropbox' or something like that. So it's actually local to everybody at the time they decide to look at it… It seems that there should be some concept of ownership, and the stuff I put out there at least, he shouldn't be able to delete... But I should be able to delete it off other people's machines. In fact, one time I did that. I put out a draft of something, and my sister objected to it, and so I deleted it and revised it and put out another one that was somewhat sanitized. So that was a reasonably good reason for me to have control over that. But it seems a little bit weird that there should be a copy on my hard disk—even stuff that my brother put out there—that he could delete it at any time. If I want it, I really have to copy it out of Dropbox and put it someplace private.*

Reasoning about the effects of deletion (especially in a shared folder) is one key to building a more robust model of synchronization.

UNDERSTANDING THE CLOUD

How do participants reveal their understandings of these cloud services? Stories about use were a primary source for eliciting a model of how participants conceptualized the cloud. Descriptions of their collaborators' actions were also a rich source of information, as were accounts of breakdowns and aspects of the services that seemed to be magic. We used a few key questions (e.g., Can you access these files offline? What happens when you delete a file? When do you think the sync happens?) to elicit participants' understandings if the topics did not arise organically. Finally, a few participants shared tricks they had learned, e.g., to get around storage quotas or other limitations.

Sometimes the interviews themselves prompted participants to form a model of how the services work. S018 reflected:

> *I mean, as you're asking this, I'm thinking, you know what, I'm really not using Dropbox to its full capacity! I actually download my entire Document folder onto a very large USB stick… and then I copy it onto my laptop before a long trip. And that ensures that I have every document up to date, because I have no idea what I'm going to end up needing to access while I'm traveling, especially if I'm gone for any length of time.*

Thresholds in Conceptual Models

The ability to use file sync and sharing tools is substantially shaped and constrained by users' conceptual models of how these tools work. The role of models is even more important when sharing files, since each collaborator's model will shape the participants' collective use of the tools and services. The most sophisticated of the participants worked around their collaborators' misconceptions. They also performed informal experiments on the services (e.g., what happens when you delete a Google Doc you do not own?) to confirm various theories about how they work.

The interviews identified five major use cases that contribute to fully understanding cloud-based syncing and sharing services. They are (in order of increasing complexity): personal cloud repository, shared cloud repository, personal replicated store, shared replicated store, and synchronization mechanism. We make a nuanced distinction between replicated storage—the idea that local folders (and, potentially, the cloud) hold identical copies of a set of files—and synchronization, the mechanism by which the replication takes place, versions are created, and conflicts are reconciled (or reflected to the users).

These use cases identify a basic conceptual framework, shown in Table 2. Assimilating the elements of this framework enables a user to have a truly robust understanding of file sync and sharing tools, and the confidence to use them in the face of imperfect bandwidth availability, storage limitations, and collaborators with less complete understandings of the underlying processes.

use cases	concepts	user actions
cloud repository	ubiquitous access	transfers files to the cloud and accesses them from any device
shared cloud repository	synchronous access	edits shared content in the cloud
personal replicated store	disconnected access	edits content on any device, even offline
shared replicated store	deletion	understands how own actions affect content on others' devices
synchronization mechanism	sync triggers; resolving conflict	Ensures sync completion, avoids conflicts

Table 2. Conceptual framework for user-facing cloud storage services

For each use case, we have identified example *concepts* and indicative *user actions* from the interview data: concepts are the underlying principles that participants must understand about the use case and user actions demonstrate that they know how to apply them. For example, if participants understand ubiquitous access they will be able to maintain a persistent store in the cloud, and will be able to transfer files among devices. Each successive concept represents an increasingly complex threshold of understanding. We populate each level of this framework with specific examples that arose during the interviews, illustrating the increasing levels of complexity and nuance.

Cloud Repository

To illustrate this level, we look at a participant who has a fairly good understanding of the role a cloud repository might play for file transfer and manual backup, but does not understand file synchronization. S051 is an accountant who uses a service called FilesAnywhere to transfer files to a colleague (but not vice-versa), to back up important work-related files, and for ubiquitous access. She adopted the service because she thought a paid service would give her better security, because her Quickbooks files were getting too large to mail to her colleague, and because it was easier to use a cloud store than a USB stick. When prompted to tell us how FilesAnywhere works, S051 said:

> *I'll save the backup file, and then copy it from my laptop over to them… And then pretty much I can sign in from anywhere I want to. Anywhere I can get a browser…Then I can create a link that I can send it to anybody…and that link will last for however many days I set it.*

Shared Cloud Repository

All participants understood the idea of synchronous access to a shared cloud repository, e.g., Google Docs, as long as they used the Google Docs applications to access the files. Far fewer participants understood Google Docs as a more general cloud repository where they could upload and store other file types. This frequently led to Google Docs being used in task-driven way: "*it's almost in the moment, as soon as that day is gone or that moment is gone, I don't care about the files anymore.*" [S014]

Concurrent access helped participants understand shared cloud repositories; many of them had witnessed collaborators' keystrokes as they concurrently updated a file (although the synchronous access had a bit of a magic quality to it). S006 said, "*I just assume that the version of the file that's being maintained by Google gets updated and saved and that becomes the new document that everybody's working from.… I've actually been on a Google Doc when I can actually see somebody typing something. So, it happens in real time, but I don't know how they make that work.*"

Personal Replicated Store

Many of the early adopters of software like Dropbox or Evernote have good conceptual models of where the files are replicated and whether they have offline access to them. For example, S052 explained the advantage of having a receipt in Dropbox: "*You stick [the FedEx receipt] in Dropbox and you've got it at home as well, so that when the customers phone you up at 3 o'clock on a Saturday afternoon, asking for something, … you've got the waybill.*"

Shared Replicated Store

Sharing a replicated store implies that each collaborator has a copy of the files and folders on his or her computer(s) and shares the local content and folder structure. Yet this experience is only seamless if the collaborators know what to expect from each other and understand the effects of their actions on the shared store. For example, S007 said, "*In a sense, the beauty of Dropbox is that it is a local copy and it is a remote copy. If it's your data and not shared. But if you're sharing it with anyone, the fact that they have deletion rights over it is a problem.*"

Synchronization Mechanism

Fewer people fully understood the synchronization mechanism that lies behind a replicated store. Participants with a good conceptual model of how a client like Dropbox or LiveMesh works could use it more effectively. While users with incomplete or inaccurate models of file synchronization could still use it, they were more apt to fall victim to its complexity and lose files (either via misunderstood version conflicts or via propagated deletes). Or they may use it inefficiently, spending extra time managing files by hand, making shadow copies, or worrying about whether files have synced. For example, S078 speculated about file conflicts he noticed:

> *I think sometimes, it might be a bug or something, where the syncing isn't correct, and it adds these extra copies… It's rare, but sometimes it does come up, and which case, I just look at the file and see if there's anything noticeably different between the two versions and most of the time I just kill the version with the parentheses and just do a delete… In that case [when it affects an entire folder], I don't really know which one is good, so I check a few and I just do a leap of faith and delete all the extra ones.*

One technique for developing a good conceptual model is to perform simple experiments with the software. For example, S092 was able to manage both bandwidth and storage limitations by experimenting with Dropbox's selective sync capability. Using this option, he discovered that specified folders remain in the cloud repository but no longer sync on a specified device: "*We haven't really done any planned cleanup, or planned work to address quota issues. ...From my perspective, the problem went away when I got selective sync working*."

Conceptual Model Discrepancies

Some participants were conscious of discrepancies in their own conceptual models or those of their collaborators. These breakdowns were helpful in testing our framework. We illustrate breakdowns in four of the levels with quotes from the interviews; we did not encounter a breakdown for personal cloud stores.

Shared Cloud Repository.

S032 helps organize an annual event that features 48 hours of live podcasts on a common theme. Prior to the event, he and the other podcast hosts use a spreadsheet to negotiate their broadcast schedule. In past years, they emailed the schedule from person to person so each podcaster could sign up for time slots and pass it on to the next person.

This year, after he introduced a Google Docs spreadsheet, S032 reported that one of the less technically savvy members of the community said, 'well, how do I send it back to you?' S032 explained, "*I already have it. It's shared. You don't have to do anything.*" Apparently, his fellow podcaster did not understand that he had edited the actual shared document, not a local copy.

Personal Replicated Store

S032 creates several audio podcasts with show co-hosts less technical than he is. He finds Dropbox a fairly simple and effective way to share large audio files and supporting files, such as sound clips and listener email, with his co-hosts. He reported initially telling them, "'*Look. Just go here. Install this program. And it's done. And you'll get a little notification when there's a file.*'" When he was asked if his co-hosts knew how Dropbox worked, S032 said:

> *To an extent. I don't think it dawns on them that if they were to install it on a different computer, there would be copies of those files there. I think they just kind of think it's in the cloud…. They don't understand necessarily that even if they weren't connected to the Internet that there'd still be a folder on their computer that has those files in it.*

If the idea of replication was vague in a participant's mind when using a system like Dropbox, there was some doubt about the scope of their actions (Would all my computers be affected?) and the effects of actions (Would changes to one file result in changes to the copies?)

While most of the reported problems with delete occur in shared replicated stores, some participants told stories about unintentionally deleting their own files from other devices in an effort to manage storage limits, as S038 recounted:

> *I've intended to use a file later, but it was maybe a big file, and I decided to pull it out of my Dropbox after I copied it there because I didn't want it there locally for some reason. And then I'd get to the other computer, and say 'oh! You pulled that out. You didn't copy it. You dipped it in there and then took it back out before*

you were ready to use it.' So that hasn't been the most intuitive thing for me.

Shared Replicated Store

Sometimes participants' misconceptions hindered them from fully using a service. For example, S018 manually uploaded files through Dropbox's web interface instead of installing a client to share local files with her collaborators: "*Dropbox I access through the web. I never even knew that [a client] was an option or thought about it. The main reason was to just to be able to share files with people.*"

Perhaps the most common misconception about shared replicas emerged when collaborators attempted to manage storage space. They deleted files from their local folder, causing the files to disappear globally when it synchronized with others' folders. For example, S092's team maintained a very large, loosely curated set of shared files:

We had a significant pile of data disappear… late last year where one of our lead manager types here thought she was cleaning up and killed a huge folder I had in [a LiveMesh folder] that was mine only. It wasn't something anyone else was using… And it was in an out-of-the-way corner. But out-of-the-way corners are the places where people sweep.

Whether the replicated files were shared between a collaborative dyad, a consultant and client, or a big team, a fundamental misunderstanding of the notion of delete was a relatively common problem among collaborators. Participants had to cross the next threshold of complexity, understanding synchronization and file versioning, to work around the unwanted deletes and potential version conflicts.

Synchronization Mechanism

When does a file sync? Does it sync at regular intervals? Is it triggered by a specific action (e.g., closing a file or saving it)? Can you ever just shut your laptop and trust that the files are in a consistent state? Users talked of various experiments and experiences trying to test their theories about the synchronization mechanism itself. Sometimes they walked away with a misperception. For example, S014 revealed the limits of his understanding of Dropbox when he tried to describe when syncing occurs:

I thought I'd figured out that it only does this every time you close the file, but I could be wrong. That's my current model, though I'm only 80% thinking that I'm accurate. The instances where I've tried to test that out, that's been the case. But there's also been instances where I'm pretty sure I left the file open at home when I came into campus and opened up my laptop, it appeared to be synced. But then there were counterexamples … There's not enough transparency to show me what was really going on.

Security and Privacy

Storing files in the cloud naturally raises privacy and security concerns. Incomplete understandings of the service by the users or their collaborators, terms and conditions imposed by the service provider, or vagaries introduced by the service provider's data handling practice can make the content vulnerable to exposure. Although the boundaries of personal privacy are fluid and negotiable [13], we wondered if experience with lower privacy expectations in other types of cloud-based services (such as social media) had inured participants to potential losses of privacy or security breaches. Can users resolve the security issues raised by complex social circumstances intrinsic to sharing files? [1, 8]

We approached these issues by asking participants what they were (and were not) willing to store in the cloud with the services they were already using, as well as noting what they did (and did not) store there now. For example, if they were using Dropbox to replicate folders, we asked them if they would be willing to replicate their entire desktop (cloud storage limits aside). We also probed explicitly for concerns about who owned the cloud and possible analytic software that the service provider ran to track users' activities and stored content. In a few cases, institutional restrictions had been imposed to prohibit employees from using services like Dropbox in the first place.

Most participants had reconciled possible security issues entailed by these services in the familiar trade-off between convenience and abstract notions of vulnerability [1]. "*The convenience just outweighs the concerns*," S006 said. To some extent, this type of response was unsurprising, since the study participants are by definition early adopters of cloud storage services, and made this decision when they adopted the service. Participants often rationalized perceived gaps in security by only exposing content that they deemed harmless if it were revealed: "*Absolutely, I would only put documents in Dropbox that could show up on the front page of the [newspaper] and I wouldn't be embarrassed… I don't put grades in Dropbox.*" [S039]

Other participants found the convenience thoroughly enticing, including S065 who said, "*If I had my way, I would have the entire computer be on Dropbox*." However, many participants listed specific types of information that they would not put on Dropbox due to security concerns: financial information, grades, business confidential information, and data revealing the identity of others (especially if governed by a research study protocol). It was important for most study participants to be able to exclude these files from synced storage, not just because they consumed storage quota on the sync service, but also because they were too sensitive to be trusted to the cloud.

Participants with a deeper understanding of the services (e.g. S007, S032, and S078) compensated for the lack of specific security provisions by adding per-file encryption when it was warranted by specific circumstances, and possibly negotiated with a colleague. S078 stored his 1Password password file on Dropbox, reasoning:

So I think if anything [my password file] might be the most sensitive thing in my life. The way I rationalize it in my head is that, one, that password file is encrypted with my 1Password password, it's an encrypted file that's on Dropbox, so even if someone else gets access to that file, it might take them a while to break into that file... And 2, I also take Dropbox at its word to say that all the stuff in there is secure and encrypted... I can rationalize it away by saying that 1Password is encrypted and Dropbox is encrypted.

Instead of reasoning about (or looking for) any evidence of the security of a service provider's cloud, some participants transferred trust in the company itself to trust in security of their cloud. S006 said, "*I guess I have a certain amount of trust in Google that I guess I believe that [Google Docs is] secure*." Similarly, S014 said, "*I guess I have different perceptions… Dropbox I know very little about, but Google as a company has a much more public figure.*"

Other aspects of the service, such as its cost, were also equated with trust. Some participants reasoned that paying for a service correlated with being trustworthy or additional security provisions, without knowing if such provisions existed. For example, S051 who paid for a service called FilesAnywhere said, "*I tend to think of [FilesAnywhere] as more secure than Google Docs. I don't know if there's anything really specific I could say to that. Y'know, maybe a false perception of security [laughs]*." S106 said of iCloud, "*My belief… is that Apple's actually not mining the data in the cloud because it's not their business model.*"

Cloud services are often offered under the condition that the service provider may legitimately run analytics on users' files. These analytics are used for a variety of purposes, including serving ads, profiling customers for other reasons, or tuning service performance. Most participants were resigned to analytics, especially if the service was free; some participants distinguished between content analytics and file characteristics (e.g., file size and type). Conversely, if participants paid for the service, they expected it to be ad-free. In fact, the absence of ads was often interpreted as evidence of a lack of analytics. S106 said he had adopted Apple's iCloud because he was in part "*…looking to move to a paid relationship with a provider, as opposed to something where my data would potentially be mined for data about me and used to push ads or whatever.*"

Since about half of our participants were technically savvy enough to distinguish between the service provider and the cloud owner, we asked participants if it mattered to them who owns the cloud that they were using. For example, if they were Dropbox users, did it matter that Amazon was ultimately responsible for providing the cloud storage?

To most, this distinction was uninteresting; they had not considered it carefully. A few participants cared (as did their institutions), but again, it was another element in the convenience vs. privacy calculus: "*That's why a lot of stuff doesn't go on Dropbox… But encrypting stuff is a pain. That's just another great way to lose your data*." [S007]. Furthermore, because this aspect of the service is usually invisible, participants knew that they should in principle care, but in practice, they did not: "*I know I should care. But I know I don't know who owns any of the clouds that my stuff is in. So I feel like I don't treat it with the seriousness that it requires.*" [S038]

In the end, the participants' use of the cloud file storage services is not only contingent on their understanding of how they work, but also on their perceptions of privacy and security concerns. These concerns are independent of the model, but do pose additional factors in adoption and use.

DESIGN IMPLICATIONS

Reflecting on our thresholds of concepts needed to form an accurate conceptual model of file syncing and sharing services leads to the question of how to enable users to build better models of these services and make fuller use of capabilities offered by the cloud. How can we get users across the thresholds identified in our framework to grow their model? We summarize our design implications in terms of process transparency, interface scaffolding [6], and reconciling conceptual models.

Process Transparency

One theme that contributed to uncertainty in users' conceptual models is not having enough cues about the process of syncing. While users did notice the notification that others had added or updated files in shared folders, there were no similar notifications about when syncing started and completed. This led to uncertainty about exactly when syncing was completed or what triggered syncing. Did syncing happen on a polling or scheduled basis? Or did the user have to save or close a file to trigger an incremental sync? Without clear transparency about the syncing process, users were left with a range of theories about when syncing occurred, often limiting their understanding or trust of the syncing process. S014 expressed this concern in suggesting how to improve Dropbox, "*Definitely more transparency. I do think it's a funny thing, where, instinctively I wanna say, just hide everything from me, I don't need to care about the server or where files sit, and I shouldn't have to. I do want to know what the background process is. Tell me what time things actually sync'ed at, give me the details of the real processes.*"

For example, each time a user opens a file in Dropbox, it could notify the user when, from what device, and by whom that file was last accessed or edited. This awareness would help users confirm that they are working on the file with the most recent changes made on another device, notice if anyone else is currently working on the file (thus avoiding future conflicts), and diagnose problems if they are not opening the version of the file that they expected.

Dropbox could also warn the user of actions that disrupted syncs that were in process. For example, if a user closed the

laptop lid or disconnected from the Internet while a sync was in process, the user could be warned that the action may have disrupted a sync process before completion (even if the user would not see this warning until the next time they accessed the file). This warning would not only alert users to the possibility of unsynchronized files, but over time also help them develop a better model of when sync happens (and more importantly, when it is completed).

Visual status indicators of the sync process (e.g., in sync, actively syncing, unable to sync), both at the individual file level and at the device/operating system level could also help users understand when files are in sync and when they might need to take extra action to restore sync among files. Just as we have learned to monitor network connectivity indicators when we need to access the internet, we could check sync indicators when we want to confirm the status of our files in the cloud.

Without this kind of process transparency, cloud services can lead the user down a garden path that the user's files are magically kept in sync with no indication of the borders and limits that actually exist. When users do encounter real-world limits to syncing, they cannot reason about why they experience problems or how to avoid them in the future. This transparency also helps users build accurate conceptual models through their use over time.

Interface Scaffolding

Our interviews elicited stories of how specific user interface interactions can either provide mental scaffolding that helps nudge users towards a more accurate model or cause uncertainty or confusion in the user's mental model. For example, we already mentioned how asking users to reflect on what happens when they delete a file in a shared folder revealed aspects of their conceptual model. When we asked S059 what happens when she deletes a file from Google Docs, she replied "*I would assume, if I'm the owner, it's deleted, but if I'm not the owner, it stays*". This seems like a fine-grained distinction, but closer inspection of how the Google Docs interface works shows how they provided the interface scaffolding to form that model.

In Google Docs, if a user tries to delete a file that someone else owns, the confirmation offered is "Remove from my Documents list?" This wording precisely conveys that the user will not see this file anymore, but others with whom it is shared will still have access to it. However, if users try to delete files that they own, the pop-up dialog choices Google Docs offers are "Trash for everyone", "Choose new owner", or "Cancel". These choices make it clear that deleting a file that the user owns will remove it for everyone unless a new owner is selected. This timely user feedback provides enough scaffolding to create an accurate model when deleting shared files.

By contrast, Dropbox gives no hints about what happens when a file is deleted, other than the standard feedback provided in the file browser. In fact, several users were surprised that dragging a file out of a shared Dropbox folder is a *move* operation instead of a *copy*, as evidenced by the file icon disappearing from the Dropbox folder. S014 described this discovery:

> *She shared her final copy of the dissertation with me and the other committee members. I went to copy it onto my own machine from a Dropbox folder. I dragged it over and then it disappeared out of the shared folder, and I thought 'Ouhh, what have I done?! I've just made the file inaccessible to everybody else on the committee!' So I quickly dragged it back into the Dropbox folder, and then sort of did a copy/paste I think. Because it didn't look to me like it was still in the shared folder once I dragged it on to my own desktop. I still had this really big sense of urgency… I thought I basically screwed up the plan for 4 or 5 other people.*

Without appropriate scaffolding, there was considerably more confusion about file deletion in Dropbox than there was in Google Docs.

Reconciling Conceptual Models

Reflecting upon the problems users encountered, we realized that a core mismatch was that users were trying to apply a familiar concept (e.g., local file storage) to a new situation (the cloud). Each of the three example services (Dropbox, Google Docs, and iCloud) tries to maintain the illusion of a familiar concept while extending its capabilities to the cloud, but breakdowns occur when the realities of the cloud do not quite fit the familiar concept.

Dropbox takes advantage of a familiar local folder model. However, confusions arise when real-world conditions prevent that folder from syncing properly or when others' interactions in the folder create surprises. Thinking of sync as a mechanism that keeps a personal local folder up-to-date across all devices and people is optimistic in a way that sets users up for breakdowns when they encounter actual limitations of the cloud. To make full use of services like Dropbox or LiveMesh, users have to master all five levels of our conceptual framework.

By contrast, Google Docs avoids this complexity by evoking a web-based conceptual model that is extended to enable working with others (both asynchronously and synchronously) through a small set of Google editing applications. Consequently, users limit themselves to sharing the core Google Doc types via the associated editors, without realizing that other file types can also be stored in the Google Docs space and thus be available across devices and people. This conceptualization may contribute to limiting the way Google Docs is used (primarily for short-term tasks like carpool sign-ups rather than the more long-lived projects we observed in Dropbox). Thinking of Google Docs as a web-based application-specific syncing and sharing mechanism is pessimistic in a way that limits full use of the cloud. While users only have to master the two repository levels of our conceptual

hierarchy, they are limited both by the technology that does not support offline access and their conceptual model of it.

Apple iCloud extends the familiar applications that users work with by invisibly syncing data across other Apple devices. While this approach short-circuits the need to understand any levels of our conceptual framework, breakdowns occur if users need that synced data for other uses, like working with a non-Apple device or application. Furthermore, new features or applications would need to be written to allow this information to be shared with others. Users' understanding of iCloud (e.g., when devices sync) is thus controlled by how much each application that uses it chooses to reveal. For example, when S106 was asked about whether deletes propagate via iCloud, he said, "*I believe they'll actually be deleted [from the other devices]. But I'm not sure if that's a preference you can set or not. If there is a default, you can change away from that, I've never bothered to play with it. Or look for it.*"

When the illusions of familiarity break down, it surprises users and undermines their development of an accurate and robust conceptual model of the cloud. Perhaps what we need is an accurate conceptual model of the cloud that reflects its capabilities and limits without trying to fit into a pre-existing metaphor. If the cloud can exhibit a user experience that helps users understand its features and limits on its own terms, then breakdowns from false expectations can be avoided.

CONCLUSION

Our survey data revealed that file sync and sharing could be improved from a user perspective. Our interviews provided more specific insights into what this improvement would entail. Crucially, users' uncertainty and misconceptions limited their ability to fully take advantage of the service's features. Users needed more accurate and robust models to be able to discover and trust cloud computing services.

We identify how process transparency and interface scaffolding can help users build better conceptual models as they use cloud-based services. As more and more cloud-based end user services are offered, we expect these insights will not only improve the file sync and sharing services that we studied, but also provide a better framework for designing more satisfying user experiences with the cloud.

ACKNOWLEDGMENTS

We thank the anonymous survey and interview participants for sharing their experiences with us. We thank Carman Neustaedter, Jed Brubaker, and our anonymous reviewers for comments on earlier versions of the paper.

REFERENCES

1. Acquisti, Alessandro & Jens Grossklags, Privacy Attitudes and Privacy Behavior, in J. Camp and S. Lewis (Eds.) *The Economics of Information Security*, Kluwer, 2004, 165-178.
2. Armbrust, Michael, Armando Fox, et al., A View of Cloud Computing, *Communications of the ACM* , 53(4), April 2010, 50-58.
3. Baker, Kevin, Saul Greenberg, and Carl Gutwin, Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration, *Engineering for Human-Computer Interaction: Lecture Notes in Computer Science*, Vol. 2254, 2001, 123-139
4. Bentley, Richard, Thilo Horstmann, and Jonathan Trevor, The World Wide Web as Enabling Technology for CSCW: The Case of BSCW, *Computer Supported Cooperative Work (CSCW),* 6(2-3), 1997, 111-134.
5. Bowers, John, The work to make a network work: Studying CSCW in action, *CSCW '94*, 287-298.
6. Chalmers, Patricia A., The role of cognitive theory in human–computer interface, *Computers in Human Behavior*, 19(5), 2003, 593-607
7. Dearman, David & Jeffrey S. Pierce, "It's on my other computer!": Computing with Multiple Devices, *CHI 2008*, 1144-1153.
8. Dourish Paul & Ken Anderson, Collective information practice: Exploring privacy and security as social and cultural phenomena. *Journal of HCI,* 21(3), 2006, 319-342.
9. Horrigan, John B., Use of Cloud Computing Applications and Services, September 2008, http://www.pewinternet.org/~/media//Files/Reports/2008/PIP_Cloud.Memo.pdf.pdf
10. Johnson, Jeff & Austin Henderson, Conceptual Models: Begin by Designing What to Design, *interactions*, 9(1), 2002, 25-32.
11. Karlson, Amy K., Shamsi T. Iqbal, Brian Meyers, Gonzalo Ramos, Kathy Lee, and John C. Tang, Mobile Taskflow in Context: A Screen Shot Study of Smartphone Usage, *CHI 2010*, 2009-2018.
12. Olson, Judith S., Gary M. Olson, Marianne Storrøsten, and Mark Carter, How a group-editor changes the character of a design meeting as well as its outcome, *CSCW '92*, 91-98.
13. Palen, Leysia & Paul Dourish, Unpacking "Privacy" for a Networked World, *CHI 2003,* 129-136.
14. Rader, Emilee, Yours, Mine, and (Not) Ours: Social Influences on Group Information Repositories, *CHI 2009*, 2095-2098.
15. Sohn, Timothy, Kevin A. Li, William G. Griswold, James D. Hollan, A Diary Study of Mobile Information Needs, *CHI 2008*, 433-442.
16. Strauss, Anselm & Juliet Corbin, *Basics of Qualitative Research*, Sage Publications, 1998.