

Overhearing the Crowd: An Empirical Examination of Conversation Reuse in a Technical Support Community

Derek L. Hansen
University of Maryland
College Park, MD
dlhansen@umd.edu

ABSTRACT

This paper describes a mixed method, empirical analysis of conversation reuse in an online technical support community. I find that the same characteristics that make the conversation successful (its highly personal, immediate, and socially engaging nature) make reuse of the conversation problematic. The archived discussion and wiki are reused to satisfy an immediate need, while the ongoing conversation is reused to help learn the practice. Use of the discussion archive and wiki repository are compared, showing benefits of the decontextualized, distilled wiki content for reuse. Implications of the findings on the design of “reuser friendly” tools and strategies are discussed.

Categories and Subject Descriptors

H5.3. Information Interfaces and Presentation: Group and Organization Interfaces---Asynchronous interaction, Computer-supported cooperative work

General Terms: Design

Keywords

Online community, virtual community, conversation reuse, knowledge reuse, email list, archive, knowledge repository, organizational memory, wiki repository, genre analysis.

1. INTRODUCTION

We learn a great deal from overhearing others’ conversations. We learn who to trust and who to distrust; how to ask and how to avoid asking; what to believe and what to doubt. Many of the greatest benefits and risks of the Internet derive from our ability to “overhear” so many conversations from so many different people.

Increasingly, these conversations are getting split apart, separated from their original context, recombined, and reused for various purposes. Unfortunately, conversations are not as socially portable as designers sometimes assume. Those who have hunted for an answer in a discussion archive or tried to make sense of a mysterious post on a Facebook Friend’s Wall can attest to this.

Conversations that occur in help-based, online support communities are particularly likely to contain valuable information for eavesdroppers. Online support communities are “distributed, computer-mediated communities of practice focused on supporting the needs of individuals who are trying to make sense of, and deal with, a particular situation or knowledge domain” [9, p. 54]. Such communities may revolve around a particular medical diagnosis, a specific technology, a scientific domain, or the latest multiplayer online game.

Copyright is held by the author/owner(s).
C&T’09, June 25–27, 2009, University Park, Pennsylvania, USA.
ACM 978-1-60558-601-4/09/06.

The question and answer discussions found in these communities’ archives provide a wealth of practical knowledge to onlookers. This value, in part, explains why lurkers far outnumber the small minority of active participants in many online communities. In addition to ongoing lurkers, countless individuals stumble upon publicly available conversations when using search engines to find answers to their own questions. This suggests that for public online support communities, the most common use is reuse.

Despite the frequency with which we reuse online conversations, surprisingly little empirical research examines the phenomenon. This is unfortunate, since a deeper understanding of how people reuse online conversations can help us design technologies and social practices to help get the most out of them. This paper helps fill this gap by empirically examining online conversations from a technical support community, *css-d*, and their reuse by community members and outsiders. I use a mixed method approach blending qualitative and quantitative methods to better understand the nature of help-based conversation and its implications for reuse. I also analyze how community members and outsiders use a community archive. After presenting the empirical findings, I discuss their design implications by highlighting the problems with many existing and proposed technical and social strategies for facilitating conversation reuse.

2. Related Work

Knowledge reuse has been studied by Information Systems and CSCW researchers under the banner of knowledge management, organizational memory, group memory, and knowledge transfer (see [13] for a review). The vast majority of studies have examined these issues within a corporate environment. They consider knowledge sharing through a variety of means such as face-to-face interactions, workflow processes, and knowledge repositories. As noted by Markus [13], there are distinct knowledge reuse situations that depend on the relationship between the knowledge creators and reusers, as well as the purpose of the reuse. The design implications for each of these can differ substantially [13]. Thus, tools and social practices that facilitate effective knowledge reuse by team members (the primary focus of early organizational memory research) differ from those that support reuse by expertise-seeking novices [13].

The current study is interested in the reuse of help-based conversations occurring in online technical support communities. Such communities consist of dozens to thousands of volunteers with a similar domain of interest and a wide range of expertise. The primary activity is asking and answering questions, although these functional activities are “fundamentally entwined with social phenomenon” [8, p. 299]. Participation is highly skewed: a small percentage of experts typically answer the majority of questions that are posed by a large number of novices (e.g., [11, 12]). To help limit repeat questions, most help-based communities encourage newcomers to check discussion archives, FAQs, or community repositories before posting questions [9, 11, 12, 14].

While sending people to discussion archives can reduce the workload for answerers, it can add a significant burden to questioners for many reasons. Discussion archives often include thousands of messages, making it a challenge to find relevant messages. This is particularly true for novices [13] who don't have an overall framework of the problem space or knowledge of specialized terminology. Furthermore, novices don't have the expertise to know which answers are of high quality or poor quality and discussion archives typically include plenty of each. Thus, even when novices come across an answer that could help, they may not recognize it as such. Or, they may not have the expertise required to apply the knowledge from a similar instance to their own [13]. In the words of Ackerman and Halverson, who based their findings on an ethnographic study of telephone hotline workers, individuals must decontextualize and recontextualize knowledge [1], a complex process that is best performed by those engaged in a familiar routine and with little time lapse [3]. The study presented here examines this issue in the context of a help-based, online community context for the first time.

Few studies have quantified the amount of information reuse in online support communities. A post-question survey by Lakhani and von Hippel found that over 3/4ths of questioners at an Apache Usenet group consulted the Apache FAQs or Usenet Archives before posing their question [11]. Questioners spent a lot of time using these resources (37.7 and 29.8 average minutes respectively) compared to the minimal amount of time it took members to post an answer to their questions (less than 5 minutes in 83% of the cases). In contrast, answerers consulted the archives in only 13% of the cases to help find the answer, as opposed to 84% of the cases where they relied upon their existing knowledge. The most active answerers read nearly all messages, primarily for educational reasons [11] (e.g., they could learn from overhearing other expert members' answers). This suggests that both novices and experts benefit from overhearing the conversations of others. Studies that examine actual usage patterns of archives are lacking.

Communities have employed a variety of strategies to make their conversations more reusable. The longstanding use of FAQ documents is perhaps the most obvious. Hansen et al., discuss the use of a wiki to distill help-based conversations into a community repository used to maintain social order (e.g., keep the discussion on topic) and help newcomers find answers to frequently asked questions [10]. Other strategies help shape the conversation as it occurs. For example, many communities have a policy that email messages answering a question should keep the same subject header as the original post, thus keeping related messages grouped together into one thread. Likewise, they recommend not posting a new question in the middle of a thread focused on a different question. Such norms may help make content more reusable, but they also may disrupt the natural flow of the conversation.

Systems designers have developed several novel approaches to better reuse conversations. Ackerman's Answer Garden [2] allows organizations to use a branching network of diagnostic questions to develop a database of commonly asked questions that grow organically over time. The system helps novices hone in on answers, but it is also limited to certain types of problems. A later version, Answer Garden 2, provides access to experts and emphasizes the need to refine the original conversations [3]. Although not focused on help-based communities, Ackerman and colleagues have developed I-DIAG [4] and Akros [15], two systems that augment human efforts to distill large discussion spaces into

more coherent summaries. In a similar vein, Brewer developed the MCS system to "condense" mailing list archives by filtering, organizing, and editing prior messages and then displaying them in structured ways intended to help novices [6]. All of these approaches provide tools to help humans transform conversations into more reusable forms. While natural language processing techniques such as automatic summarization may some day perform more of this work, they are currently not sufficiently advanced.

An alternate design strategy is to leave the conversation as is, but help provide new ways for novices to find high quality, relevant messages. As discussed, keyword searching is not always effective since novices may not know what to search for or be able to identify high quality answers. Several systems, such as Slashdot, use ratings to help separate the wheat from the chaff. These can be used to provide personalized recommendations through the use of recommender systems (e.g., [17]). Alternately, visualization techniques can be used to provide context around conversations [9, 18]. This can be useful in helping newcomers locate relevant messages and understand some of the social context around a message (e.g., it is sent from a frequent poster and is the 4th of 16 messages in the thread). Such knowledge can help better determine the credibility of information.

If we are to understand which of these various strategies and tools are most appropriate to support help-based conversation reuse, we must first have a better understanding of the nature of that reuse. Is distillation necessary or is provision of context through ratings or visualizations enough to help most reusers? Who are the people that actually reuse conversations and what are their main goals? Are the reuse challenges previously identified in corporate settings applicable to reuse of online conversations in technical support communities? What unique challenges arise in help-based conversations? The current study addresses these questions for the first time through a mixed method empirical examination of *css-d*, an online technical support community of website designers.

3. Site, Data Collection, and Methods

3.1 *css-d*

This work is part of a larger 2-year empirical study [9]. The community, *css-d*, consists of website designers interested in using Cascading Style Sheets (CSS), a technology used to add style (e.g., fonts and spacing) to structured web documents (e.g., HTML or XML documents). In practice, CSS is used to simplify the creation and maintenance of websites by separating the styling aspects from the content. Thus, like many technical support communities, its purpose is to meet the immediate needs of members as they relate to the topic at hand.

The *css-d* technical infrastructure includes an email list, public and private list archives, and a community wiki. The email list began in January 2002 and included over 7,500 subscribers as of July 2006. It is highly active with an average of approximately 50 messages sent per day. Many other messages are sent to individual list members, particularly since messages reply only to the sender by default. The continued activity of the list and its well-known reputation among website designers are indicative of this community's success.

The public list archives and wiki were created approximately 6 months after the email list and are each indexed by search engines. The public archive includes all list messages. As of May 2006, the wiki included over 100,000 words spread out over 544 pages. Community members describe the wiki as a "shared repository," "user-defined FAQ," and the "collective wisdom of this list." The

wiki is updated at a rate of approximately 65 words per day and 2 new pages per week. This modest rate of growth emphasizes the wiki's ancillary role to the email list, which is the center of community action.

Members of the *css-d* community range from novices to experts. A handful of active list administrators make sure that the discussion stays friendly, on-topic, and follows the policies (e.g., trimming quoted material). There are many other long-time members who also promote existing norms. List members describe the general tone of *css-d* as friendly and professional in contrast to other website development email lists that they participate in. This is in line with the community policies that there are "no stupid questions" and a statement by the list creator that "*css-d* is meant for beginning and experienced authors both, but I'm actually more interested in helping out the beginners." List policies suggest that members search the archives or wiki before posting, but also mention that they tolerate repeat questions as long as they are not posted soon after a related one.

3.2 Data Collection and Methods

I have chosen to study a single site in depth in order to provide a well-grounded, holistic view of conversation reuse. While this approach limits the generalizability of the findings, empirical examinations such as this one are well suited to develop theory and provide insights useful to community designers [1, 8]

The dataset includes email list messages, public discussion archive server logs, wiki content, and wiki server logs. Over 90,000 email messages were available from the list's inception in January 2002 to January 2007. Email messages were analyzed from time periods before and after the public archive and wiki were in existence. Messages that reference the email archives or wiki were oversampled, as well as messages by administrators. Wiki content included several snapshots taken between April 2003 and May 2006 and recorded in the Internet Archive or by the author. The historical snapshot approach was necessary because the wiki page histories are only stored for two-week intervals. Incutio, the company that hosts the discussion archives and wiki, provided server log reports for the September 2004 through January 2007 time period. The reports were generated by Advanced Web Statistics 6.5 software package.

These data were augmented by transcriptions of 14 one-hour, semi-structured phone interviews and 9 email interactions. Interviewees included the 4 list administrators, 8 active list and wiki contributors, 4 participants that used the email list primarily (or exclusively), 3 that used the wiki primarily (or exclusively), and 4 that rarely contributed to either. Interview questions focused on community roles, common activities, perceptions of the wiki, email list, and archives, the nature of *CSS* work, use of other information resources, motivations for certain observed behaviors, and the social dynamics of the community. Interviews were conducted concurrently with content analysis and the genre analysis so that questions about interviewees' specific actions could be addressed.

Both quantitative and qualitative research methods were concurrently used to analyze the data, allowing for triangulation. The quantitative analysis tabulated participation and usage patterns. Email list message headers and server logs were used to

characterize use of the email list and discussion archives. Wiki statistics about content creation and maintenance were calculated and usage of the wiki was evaluated via server log reports. Finally, all messages sent between Jan 2003 and April 2005 that reference

the wiki (1,897) or public archive (632) were independently coded by two raters in order to determine the reasons for the reference. Both raters coded the same 500 messages that referenced the wiki (26%) and 200 messages that reference the email archives (32%). This assured that the coding scheme was sufficiently fleshed out and consistently applied. Cohen's kappa values were above .80 for all reported statistics.

The qualitative analysis had two primary goals: 1) to characterize the nature of the conversation by identifying and examining the dominant communicative genres, and 2) to understand how users perceive the benefits and challenges of conversation reuse. The first goal was met by performing a genre analysis [16] as described in the following paragraphs. The second goal was performed through content analysis of email messages, wiki pages, and transcribed interviews. A grounded theory approach was used to identify common themes related to the reuse of email conversations. Example messages and quotes were grouped together, and tentative hypotheses and categories were developed.

A genre analysis is the process by which genres and genre systems (i.e., collections of related genres) are identified and analyzed. Orlikowski and Yates define genres as "socially recognized types of communicative actions – such as memos, meetings, expense forms, training seminars – that are habitually enacted by members of a community to realize particular social purposes" [16, p. 542]. The nuanced characterization of the discussion provided by genre analysis allows us to consider the challenges of its reuse.

I used a combination of approaches to assure that the genres I identified were recognized by the community and useful as analytic tools. First, in the spirit of grounded theory, I let the data suggest the initial genres rather than import definitions from other studies. In this case, data on genres came from my reading of thousands of email messages from many time periods, as well as interviews where I asked members to describe the activities and resources. All of the email list genres were identified by the group and verified through my reading of messages. When possible, I use *css-d*'s nomenclature (e.g., "site checks") to describe the genres or genre systems. In order to verify proposed genres and estimate their prevalence, I identified them using search strings suggested by my broader reading of messages (e.g., "site check" "holy war" "ADMIN"). My reading of large blocks of messages provided additional examples that did not have the specific search terms. Once identified and characterized, I drew upon interviews and conversations about the genres in the email discussion to explore how the genres relate to the ideas of reuse.

Names and personally identifiable information are anonymized throughout. Some messages have been slightly edited for presentation purposes. Comments in square brackets are mine.

4. Findings

4.1 The Nature of Technical Support Conversation at *CSS-D*

I think the uniqueness of *css-d* is that someone is willing to answer; someone is listening. [*css-d* list member]

Designing websites can be socially isolating work. As the previous quote indicates, *css-d* provides a direct link to others who are willing to listen and provide help. Although the conversation at *css-d* is asynchronous, it is a lively, highly personal discussion focused

on meeting the immediate needs of those willing to share their questions and concerns.

Conversation at *css-d* is typical of technical support communities more generally (e.g., [11]). It consists of regular length email messages organized into relatively short threads (avg. of 3.3 messages per thread). Although threads often last for days, it is not uncommon for replies to be sent within minutes. Participation is highly skewed. During a 28-month period, approximately 50% of all messages were sent by the 157 most prolific authors (4% of all authors). The majority of subscribers never posted. A third of posters only started threads (generally by asking a single question and then never posting again), 16% only replied to others' messages, and the remaining 53% did both at least once. As is also typical, those who answer numerous questions are typically experts, while those who only ask questions are typically novices.

The genre analysis identified three conceptually unique genre systems that are enacted in the *css-d* email list: Question and Answer (Q&A), Site Checks, and Holy Wars. In addition, two genres at the individual message level were identified: Announcements and Administrative. In this paper I only describe the Q&A and Site Check genre systems, the two most common ones. Others are reported on elsewhere [10]. Each is described by focusing on the primary situations in which they are invoked, their substance (e.g., social motives, purposes, and topics being discussed), and their form (e.g., common physical and linguistic features). In practice, multiple genres are manifest in the same email message or thread. Instead of considering each genre system as a mutually exclusive classification scheme, it should be thought of as an analytical tool used to describe the communicative actions most salient to the community itself.

4.1.1 Question and Answer (Q&A) Genre System

By far, the most common communicative form is the Q&A genre system. The socially recognized purpose of the Q&A genre system is to help meet the needs of the questioner, as well as to further educate the questioner and other list members. This genre system begins when a list member (most commonly a newcomer) recognizes an information need and brings it to the community. The typical (and suggested) practice is for her to post her question in a new thread. The new thread signals to other members that there is a new question to be addressed. Later messages reply to the initial thread providing answers, hints, comments, and links to useful references, as well as commentary on other members' suggested solutions. Following is a discussion of the question asking and answering portions of the Q&A genre system.

4.1.1.1 Question Asking

Hello list,

I would like to know how I can extend a box, built with a class, to the bottom of the browser window. Right now, the box ends where the content ends. The box uses a background image that fades from somewhat dark to light, the background color matches the lightest color in the image and the background-repeat is x-only.

Have a look at [URL including a screenshot provided] if I am not making any sense (a picture tells more than...) (200+ kB image).

Regards, Joe

In this example, Joe is unable to display the structure of a website in the desired way so he asks how to accomplish a particular effect that he believes will solve his problem (i.e., extend a box when content doesn't fill it up). Requests for procedural knowledge (i.e., knowing

"how") such as this one are far more common than requests for declarative knowledge (i.e., knowing "that"). As is typical, his question is an immediate, practical one. Joe includes a brief salutation and signs his comment demonstrating the importance of social niceties at *css-d*.

Like other knowledgeable posters, Joe uses CSS terminology to describe the desired effect. Less experienced users are equally competent at describing what they want, but less likely to use correct CSS terminology to do so. After describing his desired goal in somewhat general terms, Joe uses the majority of the message to describe the nuances of his particular site.

Joe's use of a screenshot is a practice encouraged by the list policies which state: "If you're asking for help with a problem, then remember this: A description of your problem is good. A URL to a page showing your problem is much, much better. The two together are sometimes referred to as 'mythical' or 'legendary.' Be a part of the legend." This suggestion is typically followed, although it requires constant reminders by community members. Another example that underscores the difficulty of describing visual work is shown below.

Hello,

Well I did some google-searching, and some *css-d*-archive-searching but couldn't find answer to my query. A possible reason for this is because I have no idea how to describe the problem to the search engines and... so that might be why I didn't find anything.

What's happening is that until I apply 'border' to a div that contains an unordered list whose elements are floated left, the div is not the height I would expect. It's as if the div's (or maybe it's the unordered list's) padding/margin attributes are not applied until a border is defined.

I made a page to help explain what's happening. Hopefully it will be clear what the problem is and a knowledgeable member can provide me a fix. [URL to website]

Both IE and Firefox 1.0 display this differently, but the behavior I'm specifically referring to is the one seen in Firefox.

Thanks, Frank

In this example, Frank is trying to diagnose an unexplained behavior related to the spacing of material on a website. Diagnosing unexplained phenomena is a common type of problem discussed on the list. Note the difficulty Frank has in searching for an answer. The difficulty arises from the challenge of finding the right words to describe an unfamiliar visual behavior and the numerous confounding factors. Other messages also express this difficulty (e.g., "I attempted to search the archives/Wiki on this but realized I didn't know *what* exactly to search for...") or struggle to describe the problem to the community (e.g., a member who described the "weird, bouncy disco dance" that occurs when clicking on an input field).

Like Joe, Frank provides information that he believes will "help explain what's happening" so others can effectively solve his problem. As recommended, Frank has created (and linked to) a "test case" web page designed to recreate a problem using as little CSS code as possible. This assures that other members don't have to weed through pages of irrelevant CSS code to understand the problem. Frank also includes a tentative hypothesis of the reason behind the unexplained behavior ("It's as if the...").

Note that both examples result from immediate information needs and are presented in a highly contextualized manner based on the specifics of their cases. This is the norm. Questions include stories about how the person ran into the problem, what solutions they've tried or initially thought were valid, and even occasionally why it is so important that they find an answer. These details make the interaction more personal and meaningful, as well as provide insights about the underlying information need, not simply the compromised need that is presented to the community after it has been recast in anticipation of what the community can offer [19].

4.1.1.2 Question Answering

Questions are typically answered as reply messages sent either to the individual or to the list as a whole. Some members, including at least two admins, check to see if all of the questions receive at least one reply. If they haven't, these members will provide answers or at least "point them to where they'll find more information" as one interviewee stated. Thus, nearly all questions have at least one reply. The following example is a reply message sent by a CSS expert within 20 minutes of Frank's question.

I suspect that you're suffering from some combination of "floats can stick out of their parents" and "margins collapse even between descendants and ancestors". See the following for more information on each point: *[Lists 2 URLs to web articles on "containing floats" and "uncollapsing-margins floats"]*

This example highlights the value of the human-human interaction supported by the email list. Diagnosing a complex problem is a canonical example of the type of work that human experts are adept at [5]. In this case, an expert quickly diagnosed the question that Frank had struggled to put into words.

This example also shows how resources are used at css-d to help answer questions. Here, the responder serves as an information mediator by recognizing the underlying cause of the problem and identifying resources that are helpful in overcoming that problem. For an expert familiar with the problem and resources, this activity takes little time (in this case under 20 minutes) (see [11]). Although not always sufficient to answer a question, providing links to helpful resources is highly valued and expected (e.g., "can some one please point me to a fix/hack?"; "a link to a URL or prior message would be great"). Even when the resources are not enough to solve a problem, they often help an individual reframe her question so that the individual can find an answer or rephrase the problem so that the community can answer it.

Note how the author accompanied the resource links with a short, personalized message relating them to Frank's specific question. This is the norm, partly due to the strong emphasis by administrators, one of which stated this policy when the list was still new: "Simply posting a URL as an 'answer' is also discouraged. Back up that URL with a little explanation of what the reference is about, why you posted it, and some keys to understanding the resource you're referencing. It doesn't have to be a novel; a line or two will usually suffice. But that line or two will be of enormous help to people reading your message, who may not be as expert as you are."

Like questions, answer messages are typically provided in a highly contextualized way couched in the terminology of the current problem at hand (e.g., "You are confusing ID's (#foo) with classes (.foo)..."). Some members go to extreme lengths to provide personalized answers to questions. It is not uncommon for members to download a questioner's CSS page, modify the code to fix a

problem, and then provide the modified code to the original questioner.

In addition to answering people's immediate needs, some authors like to educate the questioner about related best practices, alternate solutions, more general lessons to be learned, and additional resources. For example, after answering a specific, highly contextualized question, one author included the following line: "For your information, I diagnosed your page using my diagnostic user style sheet. See [URL]. Try it yourself..." The value of this approach was discussed by one interviewee:

I think it's more than just providing a quick answer: this property does this and this is why you need to make X-browser to Y. I think there's a lot of people on the list who really care about making sure people understand why things work the way they do, not just how to make something work.

Question answering is occasionally a collaborative endeavor. Most questions receive at least two replies, and some of the more challenging or interesting ones can spur long discussions. Although some examples of group problem solving occur, most often members simply fill in details that were missing from prior answers, post additional links to resources not mentioned, challenge or back up a particular answer, and clarify the potential problems or benefits of various proposed solutions. Sometimes these discussions turn into Holy Wars (i.e., un-resolvable debates with little practical value) that require administrators to step in, but this is not the norm (see [10] for details). On occasion the questioner is part of the collaboration as they provide additional information about their problem or clarify their needs.

4.1.2 Site Check Genre System

Another common genre system is referred to by the community as Site Checks. The socially recognized purpose of a Site Check is to provide meaningful feedback on a website, as well as to further educate the requestor of the Site Check and other list members. These threads begin when an individual solicits comments from other community members on a website that he has designed using CSS. Because the same CSS code is rendered differently by different browsers, Site Checks often include requests for members to view the site in a browser not immediately accessible. Site Check threads often include the words "site check," "site help," "browser check" or some derivation of those in the email subject line. Identifying them solely on this basis, they make up approximately 6 percent of all list messages, an average of about 3 per day. However, reply messages are often sent directly to the requestor, especially if they are about non-CSS issues.

4.1.2.1 Site Check Requests

Subject: Site check - from Windows users, please

Hi all,

I'm testing a layout for my website here: [URL]

I've included a few options to change colours and fonts in the sidebar on the right. Screenshots of how the pages look on Mac OS X in Safari are also linked from there.

If anyone on Windows systems (or any non-Mac systems basically, as Macs are all I have access to at the moment) could let me know if they notice anything out of place or significantly different from those screenshots, I'd be really grateful.

Don't pay attention to the text - also, colours in template 3 and 4 are still being worked on...

Of course, if you also want to add which fonts/colour schemes you like best, or have any suggestions for improvement whatsoever, that'd be a plus :) but the positioning and alignment of elements is really what I need to check most.

Thanks a million in advance!

Jessica

In this typical example, Jessica used a subject line indicating her need for a Site Check and a preferred platform/browser. Specific requests for browsers are common because designers often don't have all platforms available. As is also common, she provides some idea of what she would most like feedback on ("positioning and alignment of elements"). Other Site Check requests include even more explicit questions (e.g., "could someone please test the new page to see if the bottom row of navigation works?"). In addition to specific needs, Jessica asks for feedback on non-CSS issues ("fonts/colour scheme") and general "suggestions for improvement." Administrators (and other members) allow these technically "off-topic" requests to be posted as part of a Site Check, although members are encouraged to reply directly to the sender if their comments do not relate directly to the CSS aspects of the page. As is always the case, a link to the site to be tested is provided. Finally, as is often the case, Jessica thanks members in advance for their comments.

4.1.2.2 Site Check Replies

Reply messages provide reactions to the specific requests made, as well as general comments about the overall visual appeal, usability, and use of CSS on the site. They are often sent quickly after the original replies, which is part of their value. For example, some members mention that they need feedback on a particular platform since they won't have access to it until the evening. Furthermore, some people mentioned regularly scanning list messages for Site Checks since they understood their urgency.

Jessica's Site Check request received 5 public replies (and likely even more private replies). They ranged in length from one sentence to several paragraphs. The first reply sent 22 minutes after the request simply stated: "nothing changes in any of my winXP browser (opera, firefox IE6.)" followed by a signature. Another short reply message sent 5½ hours after the request stated "No Windows for testing? No problem!" followed by a URL to a web tool that provides screenshots in several browsers running on a Windows machine. The other replies include detailed comments and follow up from Jessica. The final post in the thread is a good example of a typical, detailed response.

Reply #5 of 5 Sent by Peter 7.5 hours after the original

I've looked at your page in Opera 7.50, Firefox 0.8 and IE 5.0 on win2K-pro. You've probably got it as consistent as anyone can make it, as the page holds well on screen-widths from 600 to 1600 px. Some might like the text a little larger to begin with, but none of my browsers breaks that page when resizing within the browsers, so I think it's fine as it is.

- Opera and Firefox can take that page down to a width of about 500px without creating a scrollbar, and it looks just fine.

- IE 5.0 is not centering the page, so it doesn't look the same there as in the other browsers. I wouldn't know if I couldn't compare, so the page isn't breaking or anything. It just leaves all the empty space at the right side, and breaks the page on smaller width than about 600px.

Maybe the "centering-method for IE-win" might be the thing here, but it looks good as it is too.

James has already pointed to [[URL from prior email](#)], so you shouldn't have any problems seeing that your page is doing quite well.

Regards, Peter

Like Q&A discussions, Site Checks focus on the immediate needs of the requestor and are specific to the particular situation. However, site checks also serve as fodder for discussions about best practices, common pitfalls, and novel approaches. Note how Jessica, and those listening in, learn about an unknown resource (e.g., the screen capture web tool) and potentially helpful techniques (e.g., the "centering-method for IE-win" technique briefly mentioned by Peter). Several interviewees mentioned that they learned a great deal from Site Checks because they were able to see fresh designs, as well as overhear expert critiques of them. It can be a source of design inspiration, as well as a warning sign against common problems. And of course Site Checks can be enormously helpful to the requestor. While Q&As generally address the known needs of website authors, Site Checks address the unknown needs that are recognized by experts but not the original designer.

4.1.3 Summary

Discussion at [css-d](#) is typical of many technical support communities with its skewed participation patterns and short threads. The detailed look at the Q&A and Site Check genre systems identified several important characteristics that have implications for their reuse:

- The discussion is shaped by the immediate and highly personalized information needs of requestors. As a result, most content consists of details about particular problems or websites.
- Comments, techniques, and resources of general interest, as well as social exchanges are mixed in with the personalized discussions for the benefit of requestors and those listening in.
- Questions and replies often include links to outside resources (e.g., "test cases," example CSS code, screenshots, and articles), many of which are temporary. This is an encouraged practice because of the challenges associated with effectively describing visual components so integral to website design.

Although I have discussed the major shared communicative genres in this section, there are a number of specific activities that happen within the constraints of those genres. As the prior examples show, members express gratitude (e.g., through "thank you's), provide favors (e.g., Site Checks), show off their websites and CSS knowledge, and discuss best practices. They also rant about the poor support of CSS in browsers and reminisce about the days when it "took guts to go without tables." However, they rarely discuss personal issues unrelated to website design. Several members also regularly use humor – especially humor related to website design and CSS.

4.2 Conversation Reuse at [css-d](#)

The prior section provided a nuanced description of the interaction between information requestors and providers. In this section I describe how people reuse those conversations for their own benefit. I rely on three primary sources of data including interviews, list messages that reference the list archives, and web log statistics from the list archive and wiki (for comparison).

4.2.1 Primary Reasons for Reuse

Individuals reuse the list conversations for two primary reasons. One reason is to gain exposure to new CSS techniques, design

inspiration, tools, and problem solving strategies. In other words, to learn the practice of CSS web design. One long-time member described how he continued to participate at *css-d* because the “people on it are some of the very best as in staying up to date...so you can kind of expect that if you just hang around and read post you’ll just learn a lot.” Information behavior researchers call this “orienting information seeking” where individuals monitor an “information neighborhood” surrounding an ongoing interest or concern [7]. Several well-known CSS experts joined the list as novices and attribute their current knowledge to overhearing others’ problems and then trying to answer them, providing a “test set” of real-world problems. Other members, including intermittent posters and lurkers, mentioned reading every message to learn from others’ experiences, particularly posts by knowledgeable individuals. Interestingly, these members mentioned browsing the wiki for similar reasons, but not the discussion archives (although they did use them to search for answers to their own questions as described below).

The other primary reason people reuse the conversation is to find answers to specific questions (e.g., finding a good webpage template, finding a hack around a browser problem, learning how to create a CSS-based menu). This is primarily done via keyword searching of the discussion archives or the distilled version of the conversation found on the wiki. Information behavior researchers call this “practical information seeking” that is “aimed at findings specific answers to discrete information needs, often operationalized as specific questions” [7]. One interviewee mentioned that he rarely posts since he finds almost all of the answers to his questions using the archive and wiki. Other list members also described using the archive to help answer their own questions without bothering list members. Those who answer others’ questions also mentioned using the list archives to find prior posts that could be used to answer someone else’s question, although they describe this as only rarely being necessary. Typically, when this is done answerers use their knowledge of the author (who may be themselves) and time frame to locate a particular message they remember. The specificity of the search phrases that lead people from search engines to the archive and wiki suggests that many, and likely most, reusers of those resources are searching for answers to questions.

4.2.2 Use of Discussion Archive vs Wiki

In this section I compare the use of the raw conversation archives and the distilled wiki repository. The focus is primarily on identifying the limitations and benefits of reusing conversation, not wiki repositories. However, differences in their usage often point to limitations and benefits of using the raw archives compared to a more distilled, summarized repository. See [9] for a detailed description of the wiki.

One way to compare the use of these resources is to compare how and why they are referenced in the email conversation itself. An analysis of all list messages that reference the wiki or email archives between January 2003 and April 2005 revealed several interesting findings. The wiki was referenced 5 times more often than the list archives (1,511 times vs 302 times), suggesting that it was more valuable as a reusable resource. These numbers represent only a small percentage (4.0% combined) of all 45,580 messages sent during that time period. This may be lower than other technical support communities due to the strong welcoming atmosphere, where newcomers are encouraged, but not expected to search these resources before posting.

The most common reason to reference the wiki and archives was to help answer a question or educate others more generally (61% of all messages that reference the wiki and 52% of those that reference the archives). The second most common reason was to help ask a question or justify a post (16% of all wiki emails and 20% of all archive messages). People also referenced the archive in order to complain about its usability (particularly its search feature which has a default “OR” Boolean operator instead of an “AND”) (17%), discuss its proper use (8%, $k < .80$), give thanks (4%), refer to a prior thread to continue the conversation (4%), and very rarely to discuss administrative issues such as telling others to search before posting or discouraging holy wars. The wiki was referenced far more often to deal with administrative issues, particularly avoiding Holy Wars [10] and less often to discuss usability issues.

Survey logs provide another important data point when comparing the use of these resources. They show that the archives are visited 3 times more often than the wiki (8 million times versus 2.5 million times between October 2004 and January 2007). The high numbers emphasize that the most common use of the conversations at *css-d* is actually after-the-fact reuse. For example, in July 2006 there were approximately 190,000 unique visitors with nearly 790,000 page visits. In other words, 25 times more individuals accessed the discussion as a secondary resource than there were members of the list.

Visitors came to the archives primarily from search engines (71%), sometimes from bookmarks or direct access (26%), and rarely from external web sites (3%). This pattern has been stable for the entire time period. In contrast, only 41% of wiki visitors came from search engines, 39% from bookmarks or direct access (including following links from email clients), and 20% from external web sites. The ability to update content on wiki pages that use a persistent URL contributes to the higher percentage of visitors that come from external websites (particularly given that many of them come to specific wiki pages rather than the home page). Also, the decontextualized, summarized content on the wiki may be more applicable to multiple people, making it more useful to link to from other pages than individual-specific Q&A.

Although most archive visits (76%) lasted less than 30 seconds, there were over 10,000 visits that stuck around for over 30 minutes. The wiki also showed a high percentage of visits under 30 seconds (79%). One likely explanation for this is that people are searching for specific things (i.e., an answer to a problem) when they end up accessing these resources and are able to quickly determine if the content is relevant or not. Fewer visitors come to browse as described in the prior section.

Surprisingly, in July 2006 over 66,000 of the 77,669 messages in the archive at that time were accessed (i.e., 89% of all *css-d* messages). If the log reporting software missed a web crawler, this number could be inflated. But the fact that there were 150,000 unique search phrases using 17,000 unique keywords suggests that the numbers are reasonable. The 1,000 most frequently accessed messages were just as likely to be old messages as new ones. For example, the median message rank (with 1 being the oldest message) was 39,921, which is close to 38,835 - the expected median rank of a randomly selected subset of the entire corpus based on a uniform distribution (i.e., half of the total number of messages in the corpus). Likewise, the 25th percentile and 75th percentile rank messages (23,709 and 53,753 respectively) are similar to their expected values (19,167 and 57,501). This has several important implications. First, if

