

Exploring the Relationship between Personal and Public Annotations

Catherine C. Marshall
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052 USA
cathymar@microsoft.com

A.J. Bernheim Brush
Dept. of Computer Science and Engineering
University of Washington
Seattle, WA 98195
ajb@cs.washington.edu

ABSTRACT

Today people typically read and annotate printed documents even if they are obtained from electronic sources like digital libraries. If there is a reason for them to share these personal annotations online, they must re-enter them. Given the advent of better computer support for reading and annotation, including tablet interfaces, will people ever share their personal digital ink annotations as is, or will they make substantial changes to them? What can we do to anticipate and support the transition from personal to public annotations? To investigate these questions, we performed a study to characterize and compare students' personal annotations as they read assigned papers with those they shared with each other using an online system. By analyzing over 1,700 annotations, we confirmed three hypotheses: (1) only a small fraction of annotations made while reading are directly related to those shared in discussion; (2) some types of annotations – those that consist of anchors in the text coupled with margin notes – are more apt to be the basis of public commentary than other types of annotations; and (3) personal annotations undergo dramatic changes when they are shared in discussion, both in content and in how they are anchored to the source document. We then use these findings to explore ways to support the transition from personal to public annotations.

Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital Libraries – *User issues*; H.5.2 [Information Interfaces and Presentation]: User interfaces – *Evaluation/ methodology*; H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces – *Computer supported cooperative work*

General Terms

Design, Documentation, Human Factors, Performance

Keywords

Annotation, collaboration, education, reading, study, on-line discussion, annotation system design, digital library use

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JCDL'04, June 7–11, 2004, Tucson, Arizona, USA.

Copyright 2004 ACM 1-58113-832-6/04/0006...\$5.00.

1. INTRODUCTION

Annotations on published source materials acquired from digital libraries form the basis for many subsequent collaborative activities in the classroom and in the workplace. Among these activities are online discussion (e.g. [4,5,16,24]), information brokering (e.g. [19]), and group interpretation of collected materials (e.g. [6,22]).

Past studies have shown that at least some of these shared annotations are based on personal annotations [23]. In this paper, we take an approach that enables us to more specifically characterize the relationship between the personal annotations people make while they are reading, and the annotations they share with each other when they are discussing the same materials online.

Today, in spite of the fact that much reading material is acquired from electronic resources, readers make most of their personal annotations on paper, whether they are simply responding to their reading, planning for future in-class participation, or highlighting a passage for use in future activities like writing [18]. Paper provides readers with the appropriate affordances for this sort of active engagement with a document [21].

The advent of better support for reading on a screen such as that offered by tablet computers and electronic books reopens questions about the relationship between personal and shared annotations [20]. Because personal annotations may be recorded as freeform digital ink on these current and next-generation platforms, it will be much easier to share the annotations directly.

We are left then with a central question: Can we anticipate – and potentially support – the ways in which personal annotations contribute to collaborative activities and the transitions they undergo as they are shared? Understanding the relationship between personal and shared annotations will help guide our own future design efforts, and may inform the design efforts of others working with annotation infrastructures, standards, or user interfaces.

To this end, we have performed a study to track and compare the personal annotations students made while they were reading (e.g. Figure 1a) and the corresponding annotations they contributed to online discussions of the same set of documents (e.g. Figure 1b). By analyzing over 1,700 annotations we found that only a small fraction of personal annotations were made public online and the annotations that were shared underwent dramatic changes both in content and how they were anchored to the document.

We will begin by discussing related work. Then we describe the study and the data we collected. Finally we report our findings

to a common problem.

The problem is that computer technologies are often developed and deployed, and only later is it realized that they poorly support human values. Did designers of the browser technologies, for example, anticipate that companies engaged in e-commerce would use the browser to collect (and share) information about the buying habits of individual Web-users, or track user's mouse movements? Probably not. Nor did many designers anticipate that some companies would use computer technologies to monitor the content of each worker's e-mail, let alone the number of minutes spent on-line each day or month. In other words, technologies can run roughshod over such values as privacy, autonomy, accountability, access, freedom from bias, informed consent, trust, and the right to property.

In response to this problem a multi-disciplinary approach has recently emerged, called Value-Sensitive Design. This approach seeks to design technology that accounts for human values in a principled and comprehensive manner throughout the design process.

In this paper, we first describe the Value-Sensitive Design approach. Then we bring this approach to

*be true, but many of the
no weren't highly usable and
really follow laws*

Figure 1a. A personal annotation on hardcopy of a digital document.

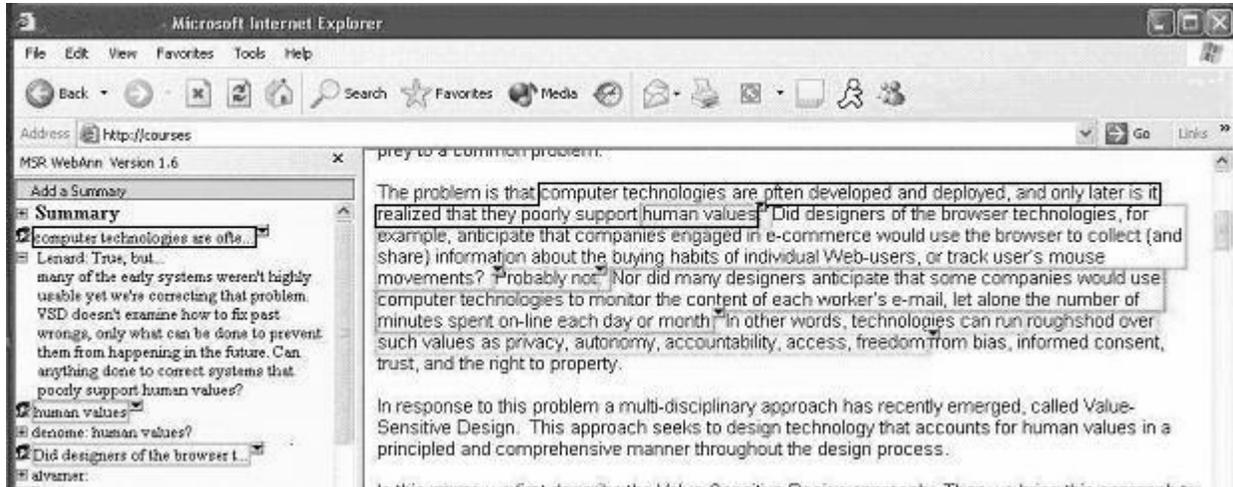


Figure 1b. The annotation's public online counterpart in WebAnn

and use them to discuss design implications and avenues for future research.

2. RELATED WORK

Three areas of previous research on annotation motivate our current study: (1) studies of personal annotation and the effect of personal annotations on subsequent work; (2) prototypes that support online discussions anchored within documents; and (3) tools and studies that address the re-use of personal annotations on digital documents.

Our current study draws on past studies of reading and readers' annotations; much of this work necessarily involves annotations on paper, since paper provides readers with a convenient, malleable surface for marking as well as the ready mobility that is a central requirement for reading and reviewing [1,17,18]. Furthermore, annotation is an essentially unselfconscious activity and many computer-based annotation tools interrupt the broader task [11,12]. Often personal annotations are shared only serendipitously, when a marked-up document or used book finds its way into other hands [12]. Past studies have also shown that personal annotations suffer crises of intelligibility: they often do not retain their meaning over time and beyond short-term tasks, even to their authors [14]. We add to this body of work by providing a detailed look at the characteristics of personal annotations and a set of categories by which they may be classified.

While past prototyping efforts have focused on collaborative annotations that serve a variety of functions, we focus on making our study results applicable to work on online document-centered discussions. There are a growing number of such tools for diverse applications including education [3,7,9], scholarly discourse [19,24], and scientific or business collaboration [5,16]. In this work we examine how the collaborative annotations supported by these tools are related to the personal annotations people make while they are reading.

Finally, we build on tools and studies that address the manipulation and re-use of personal annotations on digital documents; much of this work seeks to answer the question of how users can derive value from personal annotations made with freeform digital ink, although the work by Decurtins et al. investigates the use of digital annotations in a physical setting [8]. In [20], Schilit et al. discuss facilities including collecting annotations and retrieving related reading based on the content of annotated text. In [15], Marshall et al. discuss how personal annotations may be brought into collaborative situations. Shipman et al. present an algorithm for identifying high value annotations for use in subsequent activities [23]. Golovchinsky and Denoue [10], as well as Barger and Moscovich [2], address the issue of manipulating the digital ink itself, with the aim of tying markup to the underlying text to move it and tidy it to correspond to changes in the document display. To build on this body of past work, our discussion of the design implications of this study looks specifically at easing the transition between personal and public

annotations, including the manipulation and recognition of digital ink.

3. STUDY DESCRIPTION

Our study brings together three sources of data: semi-structured interviews with the study participants, a collection of their annotated readings on paper; and their corresponding contributions to online discussions and summaries. The interviews helped us understand the students' reading practices and gave us a general picture of how they annotated, both as they read and as they participated in online discussions. The collected annotations allowed us to investigate the relationship between personal and shared annotations in further detail.

The study participants were graduate students enrolled in a Human-Computer Interaction seminar. The 11 students represented disciplinary backgrounds that included computer science, medicine, and library science. Each week they read three selected papers on a particular topic; the papers were all provided on-line, on a class web site. Before the class met and discussed the papers face-to-face each student was required to complete two different tasks. First, they were asked to write a brief shared online summary of every paper they read. Second, they were asked to participate in online discussions of each paper.

The students used WebAnn (described below) to share summaries of their readings and conduct online discussions for 4 weeks during the 10 week quarter; the other weeks they used another threaded discussion tool, EPost [9]. A comparison of the students' online discussions in WebAnn and EPost can be found in [3]. However, for this study we draw our data from the four weeks that they used WebAnn – weeks 4, 5, 8, and 9. During those weeks, the students read 12 different papers that formed the basis for our analysis.

We selected the weeks that the students used WebAnn for two reasons: (1) WebAnn is designed to support within-document annotation, so the students' online contributions would be more readily compared to their annotations on paper; and (2) the students saw no advantage in using WebAnn's personal annotation facility, which would have enabled them to easily share their comments without retyping them. We wondered if we fully understood why.

At the conclusion of the course, we collected 49 readings on paper that the students had saved. We also had access to the students' online summaries and discussions in WebAnn. We received annotated papers from 7 of the 11 students; 2 students did not give us their papers, and 2 others did not annotate the papers as they read.

3.1 WebAnn

Students shared public online annotations using WebAnn, an extension to Microsoft Internet Explorer that supports annotating text on web pages. Figure 1b shows WebAnn. The web page being annotated is displayed in the right pane of the browser and the annotations and any replies to them are shown in a separate index pane on the left. To use WebAnn for online discussion, the students annotated HTML versions of the readings for that week.

To create an annotation in WebAnn, a student selects text in the online paper to comment on. The comment is displayed in the index pane and the selected text becomes the annotation's anchor and is outlined in the student's unique color. Thus WebAnn

Table 1. Examples of coding for personal annotations

Annotation Example	Type	Anchor Type	Content
<u>Addressing the problem. To avoid the pitfall of supporting rational "myths," learn how work is actually done. Using tailorable systems is a</u>	Underline (anchor-only)	sentence	None
✓ 7. Failure of intuition. especially poor for multi-	Mark (content-only)	heading	✓
I expect a dramatic and their g devices in use, and GUIs, implies that we are	Anchor + Content (compound)	phrase	And their GUIs

annotations are comparable to a "highlight and margin note" style of commenting with annotations linked to their context. Other students can reply to WebAnn annotations, creating a threaded discussion anchored to a particular place in the text.

To add their summaries of a reading, the students used an "Add a Summary" button at the top of the index pane, shown in Figure 1b. The summaries can be read by expanding the "Summary" heading in the index pane.

3.2 Coding method

To analyze the students' annotations, we looked at three elements of an annotation: its type, its anchor, and its content, which is any associated writing or symbols in the margin. Table 1 contains examples of coding results for several personal annotations.

To ensure the integrity of our coding, we each coded the annotations independently then compared them to find and normalize controversial codings and correspondences between personal and shared annotations. In most cases, our codings agreed at the outset. In more difficult cases, we developed rules that we could apply to resolve subsequent ambiguities. In the end, we felt like we had developed workable coding heuristics.

Annotation Type

Table 2 lists the values we used to code an annotation's type. We have divided them into three broad categories: *anchor-only*, *content-only*, and *compound*. Compound annotations consist of more than one constituent type, for example an underline and note (see Table 1).

For compound annotations we recorded specific details about the annotation's constituent types. However, to help uncover patterns we have collapsed the many variations of compound annotations into three subcategories (see Table 2, Compound row). The first, *anchor + content*, includes two important categories of common annotations, an anchored note and an anchored symbol such as an asterisk. The second, *complex anchors*, is relatively uncommon and includes annotations such as underlines with embedded circles. The third, *complex content*, we included for completeness;

Table 2. The annotation type categories

Annotation Type Categories	Annotation Types
<i>Anchor Only</i>	Underline, Highlight, Circle, Margin Bar
<i>Content Only</i>	Note, Mark (e.g. *), Other (e.g. doodles)
<i>Compound</i>	Anchor + content, complex anchor, complex content

only one instance of this category occurred in the data.

Shared annotations in WebAnn always were coded as the compound subtype of *anchor+content* since they consisted of a highlight and associated note.

Annotation Anchors

In our anchor coding scheme, we recorded *extent* (selection length) and the *structural relationship* of the anchor to the text by using the following categories: word, term, phrase, multiphrase, sentence, multi-sentence, paragraph, list, heading, document, and unanchored.

Coding correspondences

After coding all the annotations on a reading, we identified which shared annotations corresponded to personal annotations. When personal and shared annotations matched, we characterized changes to the annotation's content and anchor. The values we used for coding changes are described in Tables 6 and 7.

For many of the content-bearing annotations, determining this relationship was easy. For example, one student's personal annotation reads, "similar to cue cat; failed" and the student's corresponding comment in WebAnn reads:

"This sounds similar to the cuecat device distributed by Radio Shack free-of-charge. More information about it can be found here [URL]. As I recall, it hasn't been widely successful but some users really like it."

In the case of anchor-only annotations such as underlines, the correspondence is more difficult to determine. If the student's anchor in the WebAnn commentary closely matched the student's anchor on paper, we adjudged it a correspondence, although the relationship is weaker.

We also coded which annotations contributed to the student's online summary. It was sometimes challenging to determine whether the student used an annotation in their summary, so we were careful to be conservative.

4. FINDINGS

In this section we discuss the students' self-described reading and annotation practices coupled with our analysis of the coded personal and shared annotations. A brief account of this study and an overview of part of the data from early on in our analysis can be found in [13].

4.1 Reading and Annotation Practices

To provide context for the detailed annotation analysis, we interviewed the students about their reading and annotation practices. The interviews focused on any changes that the WebAnn discussion assignment might have brought about in the students' personal annotation habits. We asked the students to bring a few of their readings with them to the interviews so a portion of the interview could be artifact-centered.

Although the readings were available online and WebAnn allowed the students to add their own personal annotations, the students each chose to read on paper. The students' reasons for reading on paper align with the body of research in this area [1,17,18]. Some simply expressed a preference for reading on paper: "I always read on paper. I just can't read on the monitor." Often the students read their assignments away from their computers. Even the two students who did not annotate read on paper. The printed copies also enabled the students to save their annotations; one student

said she saved her readings because she's "taken the time to write stuff on them."

The students described reading the documents quickly – albeit more carefully than usual – so they could complete the online discussion assignment by the weekly deadlines. Those who annotated mostly did so as they read; they would then revisit the paper while they were completing the WebAnn assignment.

Not surprisingly, the assigned task affected the students' annotation practices in several ways. First, some students changed the form of their annotations. One student who described himself as a highlighter, not a note-taker, said that he began jotting notes in the margins: "just little short things that would remind [him] what to say." Second, some students annotated with an eye toward the interests of others in the class. One student distinguished between notes she made for herself about the technical details of the reading, and the notes she made for others on discussion points. Another reported that she actively identified topics the others found provocative. Finally, some students annotated in anticipation of what they would actually contribute. A student reported that she would "write on the paper what [she] would write in the annotation online."

Yet the students' personal markings still reflected the unselfconscious engagement characteristic to reading-driven annotations. After describing potential changes in her annotation habits, one student told us, "I don't know why I wrote some of the things I did." When another student looked over a document she brought with her to the interview, she was surprised at the number of underlines she had made while she was reading.

In addition to the assignment itself affecting practice, specific aspects of the system also changed the students' annotations. WebAnn's within-document annotations require its users to identify an anchor for every note. The students' normal underlining and highlighting habits translated fairly readily into WebAnn anchors. One student said, "I could look at the paper and see exactly where I had made my comments." However several of the students reported problems anchoring more general comments or comments that referred to longer passages: "There were one or two places where I wanted to type a comment on a paragraph, but didn't want to highlight the whole paragraph."

The results of these interviews allowed us to interpret the detailed data we collected for the analysis of the characteristics of the students' personal annotations and the specific correspondences between personal and public annotations online. In the next sections, we present this analysis.

4.2 Characterizing Annotations

Table 3 summarizes the number of personal and shared

Table 3. Overview of materials collected. Each week, three readings were assigned.

Week	Personal annotations (on paper)	Public annotations (online)	Number of papers tabulated
4	613 (86%)	97 (14%)	17
5	236 (79%)	64 (21%)	12
8	521 (89%)	64 (11%)	13
9	165 (86%)	27 (14%)	7
Totals	1535	252	49

annotations for each weekly assignment. While the number of students reading the papers and participating in the online discussion diminished as the term progressed and they fulfilled the course requirements, the proportion of personal and shared annotations remained relatively constant throughout the term. It is clear that the students made far more annotations while they were reading than they shared with each other. Since the students had reported during the interviews that their personal annotations were task-driven, we expected some to be related to those they shared and the summaries they wrote. Characterizing the individual annotations allows us to investigate patterns in which personal annotations were used collaboratively, and which were winnowed.

Personal Annotations

Table 4 characterizes the use of personal annotation types. The data presented in Table 4 suggest that the vast majority of personal annotations are anchor-only markings. That underlines predominate over highlights is not surprising, since the students are also writing notes (sometimes on separate sheets of paper), and annotators tend to use the writing implement that's in-hand [12]. Far fewer of the markings – well under 20% of the total annotations – are either content on its own or anchored content, what we traditionally think of as marginalia. These proportions are interestingly similar to those in a study reported by Bargeron and Moscovich, despite the fact that their task and participants were quite different [2].

Personal annotations were more frequently anchored at the sub-sentence level (691, or 45%) than at the sentence level (402, or 26.2%) or to longer segments of text (263, or 17.1%). We will revisit this finding when we discuss anchoring patterns in shared annotations.

Public Annotations

As we described earlier, WebAnn constrains the representation of within-document annotations to all be instances of *anchor+content*. Thus the important annotation type distinction for the 252 public annotations online is whether they represent a shared summary (48 or 19%) or are part of the online discussions (204 or 81%). Note that of the 204, 121 are original comments and 83 are replies.

While personal annotations were more commonly anchored at the

Table 4. Overview of personal annotation types

Annotation Type	Number	Frequency
Anchor only	1262	82.2%
Underline	842	54.9%
Highlight	250	16.3%
Circle	140	9.1%
Margin bar	30	2.0%
Content only	120	7.8%
Note	83	5.4%
Mark (eg. *)	21	1.4%
Other (eg. doodles)	16	1.1%
Compound	153	10.0%
Anchor + content	138	9.0%
Complex anchor	14	0.9%
Complex content	1	0.1%
Total	1535	100%

sub-sentence level (as reported above), shared annotations were linked primarily to sentences (51 of 121 original comments or 42% were anchored on sentences), while only 23 of 121, or 19%, were anchored on words and phrases. A qualitative examination of these anchors shows them to be on more carefully selected portions of text. This finding foreshadows the kinds of anchor changes we found when we looked at individual relationships between personal and shared annotations.

4.3 Using personal annotations in online discussions

When we began our study, we entertained three hypotheses about the relationship between personal and public annotations:

- only a small fraction of the personal annotations that people make while reading are shared with others;
- some types of annotations are more likely to be shared than others and some annotation styles are more readily sharable;
- annotations made while reading undergo profound transitions when they are made public; both their content and the extent of their anchors are changed to make the annotations intelligible to others.

We devote a subsection to each hypothesis; we also briefly discuss the potential effects of individual annotation styles on sharing.

Small fraction of personal annotations shared

To investigate the first hypothesis, we examined how many of the annotations the students made on paper corresponded to annotations they made online. We discovered that of the 1535 personal annotations that the students made while they were reading, only 379 corresponded in any way to their shared annotations (Table 5, bottom row). Of the 379 correspondences, 259 were related to the shared summaries and 120 were used as a basis for the online discussions. Thus, a little less than a quarter of the personal annotations could be tied to the collaborative annotations, and still fewer – about 7.8% – may have been the basis for the comments the students made in discussion.

Annotation types that are shared

By examining the annotation type categories, we can get a more nuanced sense of how the personal annotations were related to the collaborative annotations shared online. Table 5 summarizes the relationship between annotation types, and how they were used by the students. The highlighted rows in Table 5 show the types of annotations most apt to be shared, *notes* and *anchor+content*. The frequencies following the numbers represent the percent of the total number of that type of annotation.

One trend evident in Table 5 is that the existence of specific content is a good predictor of whether an annotation will be the basis of comment shared in the online discussion. This finding is supported by the interview data; students referred to jotting down notes to remind themselves of what they might share online. In fact, the *anchor+content* style of annotation (from the compound category) is the most likely to make the transition from a personal annotation to a contribution to the online discussion (84 of 138 or 60.9%), followed by the *note* annotations (42 of 83, 50.6%). (See Table 5, “Total shared online” column.)

Anchor-only annotations were the least likely form the basis of shared within-document commentary (3.3% of them were re-used in this way) (Table 5, last column). While it is possible that anchor-only annotations help the students recall portions of the

Table 5. Correspondences between personal annotations (on paper) and public annotations (online). Corresponding public annotations are further broken down by online use in summaries and as commentary.

Annotation Type (on paper)	Annotations on Paper (total)	Corresponding Public Annotations Online		
		Total shared online	Number used in online summaries	Number used in online discussion
Anchor only	1262	247 (19.6%)	205 (16.2%)	42 (3.3%)
Underline	842	167 (19.8%)	135 (16.0%)	32 (3.8%)
Highlight	250	46 (18.4%)	38 (15.2%)	8 (3.2%)
Circle	140	31 (22.1%)	29 (20.7%)	2 (1.4%)
Margin bar	30	3 (10.0%)	3 (10.0%)	0 (0.0%)
Content-only	120	44 (36.7%)	21 (17.5%)	23 (19.2%)
Note	83	42 (50.6%)	20 (24%)	22 (26.5%)
Mark (e.g. *)	21	2 (9.5%)	1 (4.8%)	1 (4.8%)
Other (e.g. doodles)	16	0 (0.0%)	0 (0.0%)	0 (0.0%)
Compound	153	88 (57.5%)	33 (21.6%)	55 (35.9%)
Anchor + content	138	84 (60.9%)	29 (21.0%)	55 (39.9%)
Complex anchor	14	4 (28.6%)	4 (28.6%)	0 (0.0%)
Complex content	1	0 (0.0%)	0 (0.0%)	0 (0.0%)
Total	1535	379 (24.7%)	259 (16.9%)	120 (7.8%)

text that they want to return to and comment on, the effort of re-interpreting the text to contribute in a collaborative situation is necessarily higher. The anchor-only annotations were far more apt to form the basis for the students' summaries (16.2% of them were used this way); previous research has shown anchor-only annotations may be used to designate what the reader feels is important in the text [12].

Profound changes in content and anchors

To better characterize the changes the students made when they shared their personal annotations with each other using WebAnn, we will focus on the 120 annotations that they turned into anchored commentary (Table 5, last column).

First, we need to examine how the students changed the *content* of their personal annotations to make them intelligible to others. As defined in Table 6, we coded 5 categories of content changes: *cleaned up*, *original and more*, *cryptic to understandable*, *nothing to something*, and *unrelated*. Table 6 shows that only 8.3% of the

Table 6. Content changes that occur when annotations are shared

Type of content change	Description	Number (% of total)
Cleaned up	More or less verbatim of paper annotation	10 (8.3%)
Original and more	Include and expand on paper comment	41 (34.2%)
Cryptic to understandable	Profound change to make intelligible	16 (13.3%)
Nothing to something	Anchor-only on paper, comment online	52 (43.3%)
Unrelated content	Anchor of paper and online annotation match, content differs	1 (0.8%)
Total		120 (100%)

personal annotations were used as is, or simply tidied up; most are significantly extended when they are shared. Almost half (43.3%) are in the *nothing to something* category, meaning that the student is reminded of his or her interest by the anchor on paper, but no interpretive content or reaction was actually written down while the student read.

Anchor changes were almost as common as content changes when annotations were shared; 80% (96 out of 120) changed. Table 7 summarizes how they changed. Since sub-sentence and non-contiguous anchors are more common in personal annotations than in shared annotations, it is not surprising that the most common kind of change is to extend the anchors. In general, these changes reflect the informality of the students' personal annotations. Anchors used for sharing commentary are more precise, singling out the specific text that triggered the comment.

Stylistic differences among individuals

Individual students differed in their annotation practices. Are there aspects of individual style that may affect sharing? In the interviews, the students were able to describe their own personal annotation styles, even if they did not always recall how much

Table 7. Anchor changes that occur when annotations are shared

Type of anchor change	Description	Number (% of total)
Verbatim	Identical to paper	24 (20.0%)
Cleaned up	E.g. extend to sentence boundary	4 (3.3%)
Smaller extent	Shorter anchor online	23 (19.2%)
Greater extent	Longer anchor online	32 (26.7%)
Nothing to something	Unanchored on paper, explicit anchor online	19 (15.8%)
Moved	Different anchor online	18 (0.8%)
Total		120 (100%)

Table 8. Quantity of personal vs. shared annotations

Student	# annos (# papers)	Av. annos per paper	Annos on paper	Annos online
1	273 (7)	39	239 (88%)	34 (12%)
2	132 (2)	66	121 (92%)	11 (8%)
3	258 (8)	32.3	232 (90%)	26 (10%)
4	551 (10)	55.1	493 (89%)	58 (11%)
5	176 (9)	19.6	129 (73%)	47 (27%)
6	180 (7)	25.7	137 (76%)	43 (24%)
7	217 (6)	36.2	184 (85%)	33 (15%)

they had annotated, or why they had made a particular annotation. The readings we collected confirm that each student has a recognizable personal annotation style.

Table 8 shows that if a student produces a large number of personal annotations, it does not necessarily imply that the student will share a proportionally large number in discussion (and two students participated in the WebAnn discussions without annotating at all while they were reading). Thus frequency alone is not a good predictor of how many annotations will be shared.

Is preferred annotation type a better predictor? Will the students whose personal annotation style favored anchor+content annotations, the type most commonly shared, contribute more to the online within-document discussions? Despite the intuitive appeal of this inference, this stylistic preference doesn't necessarily increase sharing. For example, one student who averaged over 6 shared annotations per reading made about 2 anchor+content style personal annotations per reading, and another student who averaged between 5 and 6 shared annotations per reading made around 9 anchor + content style personal annotations per reading.

The existence of individual styles has design implications that we will explore in the next section, but in our study use of a particular style did not ensure increased sharing.

5. DESIGN IMPLICATIONS

If tablet computers deliver on their promises, we can expect an increasing number of people to read on the screen and annotate digital materials as they would paper documents. In fact, past studies have shown that personal annotation styles and practices translate fairly readily to tablet-based annotations [14]. Freeform digital ink will move from prototype [20] to commonplace, and the broad range of personal annotations we explored in this study will not need to cross the paper-digital divide to be used in collaborative situations. At the same time, systems that support these collaborative situations – writing, discussion, design, interpretation, and others – are becoming similarly mature and are accepted part of many educational and office information technology environments.

Given that the means of producing personal annotations on a computer are close at hand, as are the venues for sharing annotations online, we can turn our focus in future work to ways of supporting the transition between the two. Our study provides ample evidence that this support will require a more sophisticated strategy than simply adding a “Publish my annotations now!” button. Annotations on a single document may be used in multiple ways. For example, in our study, they were used to write summaries as well as being the basis for public commentary; they

may have also served to simply focus the students' attention while they read the assignments. Furthermore, depending on who they will be shared with and why, only a small number of personal annotations are used collaboratively and those that are, are edited.

We can divide the transition into two different activities. First the reader must find and review annotations that are potential candidates for sharing; when marks are plentiful, this is not a simple problem. Second, once the reader has found an annotation to share, he or she must render it suitable for sharing: it must be made intelligible.

5.1 Help find and filter annotations for re-use

Our findings show that relatively few (at best 7.8%) of the students' personal annotations translate into collaborative use. This selectivity suggests that aiding in navigation through an annotated document, and filtering for the annotations most likely to be re-used may be a successful strategy in helping people find the annotations they want to share. The effectiveness of this type of facility can be amplified with a good interface for establishing and maintaining individual annotation styles so that an annotator naturally develops useful annotation habits.

For example, one of the students used 4 out of 6 of her content-bearing personal annotations on a reading in the online discussion. All except one of these were an underline coupled with a note; if she could step through this type of annotation, ignoring her far more numerous underlines, she would be reminded of the annotations she wanted to add to the discussion.

Why navigation instead of extracting annotated passages and compiling results separately? Past research showed that simply collecting a reader's annotations in a separate Reader's Notebook for use in a group discussion was not as useful as expected [14]. In fact, readers sometimes didn't recognize their own annotations when they were taken out of context. So any find or filter functionality should present the marks in context, at least initially.

Requiring in-context display introduces some interesting alternatives: showing the document with its personal annotations side-by-side with the shared version of the document, or displaying the personal annotations likely to be shared in the collaborative context. This technique would help the annotator identify points of overlapping interest with other discussion participants, as in [15], and might be used to guide the discussion, for example, when many of the students circled the same term in the text.

Our high-level coding categories may help address the problem of sorting out the kinds of annotations people make without requiring them to adhere to a strict regimen or to maintain consistent marking conventions (e.g. assigning meaning to specific pen or highlighter colors). Using current parsing and recognition techniques (such as those explained in [23]), it would be possible to distinguish among anchor+content, content-only, and anchor-only personal annotations, or to find specific kinds of annotations, for example anchor+mark, that are predictably useful.

It is also important to remember that different activities may draw upon different types of annotations. Simply filtering according to the annotation type might help the students step through and gather one type of annotation. So, for example, anchor-only personal annotations might be helpful for producing summaries, while anchor+content personal annotations might form the basis of what the annotator shares in discussion.

5.2 Support the transition from personal to shared annotations

Once candidate annotations are identified, is there any kind of processing or user interface that will ease the annotations' transition from personal to collaborative? Which processing and user interface strategies make sense and which don't?

As our findings have shown, a substantial transition takes place: the content of the personal annotations is apt to be clarified and extended and the anchors are apt to be tidied up to more accurately reflect the scope of the comment; a shared annotation reflects far more authorial intent.

The nature of the content changes we observed suggests that there is no real shortcut that supports this aspect of the transition. Given the disparity in both length and literal content of the students' personal annotations and the ones they contributed to the online discussions, handwriting recognition strategies or digital ink editing facilities wouldn't be much help. The ability to manipulate either recognized text or the digital ink itself would only be sensible if the personal annotation and shared annotation were close in content; in our study, less than 10% were. Returning to the example, we used in the previous subsection, the student's on paper comments were brief – from 2 to 10 words long. Her online analogs varied from 32 to 150 words, a considerable extension.

More apt would be to support the changes in anchoring. A better user interface for anchoring and modifying existing anchors (tidying them up in both form and position, and knitting together non-contiguous spans) would ease this aspect of the transition. The relevant underlying technology here is that described by Golovchinsky and Denoue [10], or that described by Barger and Moscovich [2].

6. CONCLUDING REMARKS

In our study of the relationship between personal and shared annotations we have found that only a small fraction of personal annotations are made public online, that some types of personal annotations are shared more often than others, and that annotators make profound changes to annotations that they share online. These findings suggest designers of annotation systems should assist users in finding and filtering reusable annotations and support the transition from personal to shared annotations.

We have addressed the transition from personal to public annotations as if the personal annotations people will make while they are reading on tablet computers will be just like the personal annotations they make on paper. Is this a valid assumption? At least initially, there is evidence that it probably is (see, for example, [14]).

As people gain more experience reading on tablet computers, will their personal annotations change so they can be shared as is? We suspect not. Usually personal annotations reflect unselfconscious reactions to reading material, while public or shared annotations on a document reflect specific communicative intent. It is important to most readers that annotating does not interrupt reading, although we expect that in a handful of situations people may make digital ink annotations with the explicit intention of sharing them, much like an instructor currently shares her handwritten comments on student work with the student. More likely is that personal annotation habits will evolve in minor ways to take advantage of the kinds of functionality we describe.

7. ACKNOWLEDGMENTS

We would like to thank the students who so graciously allowed us to collect and analyze their readings.

8. REFERENCES

- [1] Adler, A., Gujar, A., Harrison, B.L., O'Hara, K., and Sellen, A. A Diary Study of Work-Related Reading: Design Implications for Digital Reading Devices. In *Proceedings of CHI '98* (Los Angeles, California, USA, April 18-23, 1998). ACM Press, New York, NY, 1998, 241-248.
- [2] Barger, D. and Moscovich, T. Reflowing Digital Ink Annotations. In *Proceedings of CHI '03* (Ft. Lauderdale, Florida, USA, April 5-10, 2003). ACM Press, New York, NY, 2003, 385-393.
- [3] Brush, A., Barger, D., Grudin, J., Borning, A. and Gupta, A. Supporting Interaction Outside of Class. In *Proceedings of CSCW '02* (Boulder, CO, USA, January 7-11 2002). Lawrence Erlbaum Associates, Hillsdale, New Jersey, 2002, 425-434.
- [4] Cadiz, J., Gupta, A., and Grudin, J. Using Web Annotations for Asynchronous Collaboration around Documents. In *Proceedings of CSCW '00* (Philadelphia, PA, USA, December 2-6, 2000). ACM Press, New York, NY, 2000, 309-318.
- [5] Churchill, E., Trevor, J., Bly, S., Nelson, L., and Cubranic, D. Anchored Conversations. In *Proceedings of CHI '00* (The Hague, The Netherlands, April 1-6, 2000). ACM Press, New York, NY, 2000, 454-461.
- [6] Cox, D. and Greenberg, S. Supporting Collaborative Interpretation in Distributed Groupware. In *Proceedings of CSCW '00* (Philadelphia, PA, USA, December 2-6, 2000). ACM Press, New York, NY, 2000, 289-298.
- [7] Davis, J.R. and Huttenlocher, D.P. Shared Annotation for Cooperative Learning. In *Proceedings of CSCW '95* (Bloomington, IN, USA, October 17-20, 1995). Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1995, 84-88.
- [8] Decurtins, C., Norrie, M.C., and Signer, B. Digital Annotation of Printed Documents. In *Proceedings of CIKM '03* (New Orleans, LA, USA November 3-8, 2003). ACM Press, New York, NY, 2003, 552 - 555.
- [9] EPost, UW Catalyst Toolkit, <http://www.catalyst.washington.edu/tools/>
- [10] Golovchinsky, G. and Denoue, L. Moving markup: repositioning freeform annotations. In *Proceedings of UIST '02* (Paris, France, October 27-30, 2002). ACM Press, New York, NY, 2002, 21-30.
- [11] MacKay, W.E. Is Paper Safer? The role of paper flight strips in air traffic control. *ACM Transactions on CHI*, 6, 4 (December 1999), 311 – 340.
- [12] Marshall, C.C. Annotation: from paper books to the digital library. In *Proceedings of DL '97* (Philadelphia, PA, USA, July 23-26, 1997). ACM Press, New York, NY, 1997, 131-140.
- [13] Marshall, C.C. and Brush A.J. From Personal to Shared Annotations. In *CHI '02 Extended Abstracts* (Minneapolis, MN, USA, April 20-25, 2002). ACM Press, New York, NY, 2002, 812-813.

- [14] Marshall, C.C., Price, M.N., Golovchinsky, G., and Schilit, B.N. Introducing a Digital Library Reading Appliance into a Reading Group. In *Proceedings of DL '99* (Berkeley, CA, USA, August 11-14, 1999). ACM Press, New York, NY, 1999, 77-84.
- [15] Marshall, C.C., Price, M.N., Golovchinsky, G., and Schilit, B.N. Collaborating over Portable Reading Appliances. *Personal Technologies*, 3, 1 (March, 1999), 43-53.
- [16] Microsoft Office Web Discussions, <http://office.microsoft.com/assistance/2000/WebDiscussions.aspx>
- [17] O'Hara, K. and Sellen, A. A Comparison of Reading Paper and On-Line Documents. In *Proceedings of CHI '97* (Atlanta, GA, USA, March 22-27, 1997). ACM Press, New York, NY, 1997, 335-342.
- [18] O'Hara, K., Smith, F., Newman, W., and Sellen, A. Student Readers' Use of Library Documents: Implications for Library Technologies. In *Proceedings of CHI '98* (Los Angeles, California, USA, April 18-23, 1998). ACM Press, New York, NY, 1998, 233-240.
- [19] Roescheisen, M., Mogensen, C., and Winograd, T. *Shared Web Annotations as a Platform for Third-Party Value-Added Information Providers: Architecture, Protocols, and Usage Examples*. Technical Report STAN-CS-TR-97-1582, Stanford University, Stanford, CA, April 1995.
- [20] Schilit, B.N., Golovchinsky, G., and Price, M.N. Beyond Paper: Supporting Active Reading with Free Form Digital Ink Annotations. In *Proceedings of CHI '98* (Los Angeles, California, USA, April 18-23, 1998). ACM Press, New York, NY, 1998, 249-256.
- [21] Sellen, A. and Harper, R. *The Myth of the Paperless Office*. MIT Press, Cambridge, MA, 2001.
- [22] Shipman, F., Airhart, R., Hsieh, H., Maloor, P., Moore, J.M., and Shah, D. Visual and Spatial Communication and Task Organization in the Visual Knowledge Builder. In *Proceedings of GROUP '01* (Boulder, CO, USA, September 30-October 3, 2001). ACM Press, New York, NY, 2001, 260-269.
- [23] Shipman, F.M., Price, M.N., Marshall, C.C., Golovchinsky, G., and Schilit, B.N. Identifying Useful Passages in Documents Based on Annotation Patterns. In *Proceedings of ECDL'03* (Trondheim, Norway, August 17-22, 2003). Springer Verlag, Heidelberg, Germany, 2003, 101-112.
- [24] Shum, S.B., and Sumner, T. JIME: An Interactive Journal for Interactive Media. *First Monday*, 6, 2 (February 5, 2001).