Chapter 5 --- Empirical Investigations

In the sections that follow two collaborative tools will be examined: DocReview (see §4.3) and Research Webs (see Chapter 3). DocReview is a web-based tool that allows readers of documents to become reviewers. This critical capability allows the collaborators to correct, expand, and refine the documents. DocReview is an integral part of Research Web Essays, the principal textual tool of the Research Web.

The Research Web is a customizable collaborative environment that permits the research team in a long-term, large-scale enterprise to examine an issue domain thoroughly. The Research Web (RW) has a WWW site that serves as the repository of the team's corporate memory and research results. Tools available include a basic set that includes scholarly services of an annotatable bibliography and glossary, and an augmented web page format used for research essays. It incorporates any tool that the team finds necessary to its mission, provided that tool can be made web-compatible. Research Webs are unique, and for that reason may best be examined as case studies.

Table I Definition of Terms Used in Case Studies of DocReviews

Author, authoring team: the owner(s) of a document.
Base document: the document under review in DocReview.
Comment: the contents of an annotation returned by a reader.
Effectivity: the ratio of the size of comments received to the size of the review segment or the size of the base document.
Nature of participation: the number of people participating, the number of comments received, the volume of commentary (size), and the value [subjective] of the comments received.
Notification: e-mail sent to the reviewing team whenever a comment is received.
Quality of document, type of document: a score of value to the enterprise from 0 to 5 with zero being irrelevant and five being an essential product of the enterprise.
Review segment: a fragment, or chunk, of the base document that is the focus of annotation in DocReview. Usually a paragraph, element of a list, or a graphic insert.
Reviewing team, reviewers: people asked to participate in the DocReview.
Size (of document, comment, or review segment): the number of words longer than three letters.
Social character of comments: the content of the comments coded using the Bales IPA codes (see §5.1.4.2 below)
Substantive character of comments: the content of the comments coded using the Meyers structurational argumentation codes (see §5.1.4.2 below)

5.1 Case Studies of DocReview Installations

Since 1995, over 400 DocReviews have been installed. There are (or have been) at least nineteen DocReview sites, fifteen at the University of Washington, and one each at: University of California at San Diego, Haverford College, and the University of Wisconsin. Since the software has always been offered free of charge over the Internet, it may have been installed on several other servers, but these installations are likely to be inaccessible passworded intranets. The author has hosted several DocReview

installations for other researchers to see if the system would help them in their work. Most of these installations were made at the request of people who wanted to use it as a tool to review independent documents, rather than as a part of an enterprise approximating the nature of a Research Web.

Five sets of DocReviews have been selected for detailed quantitative analysis in order to examine several propositions (see §5.1.6) that arise from the basic research questions. The basis of selection was their similarity to DocReviews that might be found in active Research Webs. These DocReviews were all under the control of a knowledgeable facilitator (the author). These 101 DocReviews contained 1929 review segments that attracted 294 comments. These comments were coded into 767 Bales codes (see Table V) and 425 Meyers argumentation codes (see Table VI). The data was mounted in a relational database to support data conditioning and analysis. Analysis was performed with a spreadsheet program.

Two of the selected sets of DocReviews were the minutes of 59 meetings. The meetings were task-oriented meetings with an attendance averaging six members, with occasional participation of others by telephone. The minutes were quite comprehensive and averaged two pages of text. DocReview was integrated into the meeting routines by directing the attendees to review the minutes on the WWW before the next meeting. At the next meeting the scribe would distribute copies of the minutes with commentary inserted inline. The scribe would then explain how the minutes were revised in light of received annotations, and the team would then approve the minutes or suggest other changes. Usually this discussion was over in two or three minutes, thus saving considerable meeting time.

One set of seven DocReviews was sections of a draft of a professional paper. The paper was divided into seven sections in order to reduce the time required for each reviewing session. In reducing the review time, the busy schedules of the reviewers could accommodate the small time slices. The reviewers were professional colleagues of the author, some of who were involved in the design of DocReview. The author found the annotations very useful and most were incorporated into the final draft of the paper.

Another set of DocReviews was 19 workshop position papers for the 1999 conference on Computer-Supported Cooperative Learning (CSCL). Reviewing position papers was seen as an excellent application of DocReview from the beginning of design. In practice it lived up to its presumed promise. Perhaps the greatest impact was not intellectual, but in opening networking channels.

The final set of DocReviews was a set of 17 documents, Research Web Essays, written for a Research Web for the issue domain of chromium (CrVI) contamination on the Hanford Nuclear Reservation. The set was quite successful in accomplishing the objective of refining the initial versions of the documents, each of which centered on one aspect of the contamination.

5.1.1 Research Questions

The major research questions and the propositions derived from those questions are:

A. How does the behavior of dialog using DocReview compare to dialog that is face-to-face?

Does the *social* character of comments in DocReview differ from comments in face-to-face dialog?

Does the *substantive* character of comments in DocReview differ from comments in face-to-face dialog?

B. How should DocReview be segmented in order to maximize the effectiveness of the participants?

For complete DocReviews

- Long base documents are ineffective relative to shorter documents.

For DocReview segments receiving comments:

- The amount of commentary received on a review segment will be directly proportional to the segment's length.

- The ratio of size of comments received to size of review segment will decline proportional to review segment size. (Long segments are less effective than short segments.)

C. How does the design of DocReview serve the research team?

Products similar to DocReview will emerge and will, by similarity, validate the design.

D. How does the quality of the document being reviewed affect the participation in the review?

Higher quality documents will attract more participation.

The nature of social commentary will vary with the type of document.

The nature of *substantive* commentary will vary with the type of document.

5.1.2 Design of Data Collection System

DocReviews have a very complete set of data. There are three types of data: text data, installation data, and transaction data. Text data includes the base document for the review and the annotations received from reviewers. The text data is contained in files on in single directory on the server (either active or archived). The text files include: the base document (basedoc.html), the HTML code for the document to be reviewed; comment files, either in a cumulative file (cummulate.html), or in files that contain the commentary on each commented review segment (point[nnn].html). Installation data is a set of parameters established by the editor at creation. Installation data includes the name and e-mail address of the sponsor, a list of e-mail addresses to be informed when annotations are received, and a descriptive title of the DocReview. The log file (log.txt) collects transactions (creation, annotation, reading and archiving) on the DocReview as they occur.

A program written by the author extracts and formats data from the files mentioned above. The program (makecsv.pl) creates several comma separated variable (.csv) files suitable for import into a database and thence to a spreadsheet program for the analysis. This program also does a word count on the base document and each of the document's review segments.

The analyst supplements two of the .csv files in order to add information that cannot be automatically extracted. A file (docrev.csv), that captures attributes of each DocReview, is augmented by including a description of the DocReview and a document type attribute

designed to indicate the degree of quality, or the degree of completeness, of the document. This attribute is entered as a number from 1 to 5 and is defined as:

- 1 -- a sketchy document designed to collect ideas or impressions before investing any more effort in the document. A brainstorming dialog.
- 2 -- a first draft of a document, designed to catch major omissions, correct big mistakes and perhaps attract other authors.
- 3 -- a working document not intended to be advanced to publication. Minutes and position papers would fit here.
- 4 -- an advanced draft meant to polish the document before release as an essay for local use, or before submission to a journal for publication.
- 5 -- a released essay, or a published article. Under review for picking up new material as time passes, or to attract rebuttals or support.

The coder modifies the comments.csv file to add both the Bales codes (Interpersonal Process Analysis) and Meyers codes (Structurational Argumentation).

5.1.3 Quantitative Descriptive Statistics

There are three levels of abstraction used in the evaluation of DocReview: the base document, the document being reviewed; the review segments, sections of the base documents that are the annotatable units; and the annotations or comments that are made by the reviewers. This section reviews the descriptive statistics for the quantitative data collected on these levels of abstractions. The unit of analysis is a count of words of four or more characters.

Base Document

The document that is prepared for DocReview is called the base document. It varies in size, and quality (the degree of development). Very large base documents are usually

broken into sections, each a DocReview, in order to allow the usually busy reviewers to complete a section at one sitting.

Data collected on the base document for a DocReview includes a word count, the document type (quality), the sponsor (author), and date of creation of the DocReview. The text of the base document is also available.

Table II Words in Base Document

Base Document Size (word count)				
Characteristic	***All***	***Type 2***	***Type3***	***Type 4***
Mean	459.26	135.61	465.47	798.82
Median	422	130	469	598
Standard Deviation	325.27	91.259	196.71	695.66
Sample Variance	105801	8328	38693	483946
Kurtosis	20.77	-0.59	2.49	5.44
Skewness	3.44	0.48	0.88	2.22
Range	2647	309	1279	2657
Minimum	10	10	140	206
Maximum	2657	309	1279	2657
Count	100	13	76	11

The sample variances are too heteroscedastistic to employ an F test, so a logarithmic transformation was attempted in hopes of reducing the sample variances. The transform also failed to reduce the heteroscedasticity to an acceptable level. Differences between document types cannot be said to exist based on the base document word count.

Review Segments

Commentary other than general comments are directed toward a fragment of the base document called a review segment. Review segments are most frequently paragraphs or list elements (bullets), but occasionally include images or entire tables. The facilitator determines the review segments. The DocReviews in this case study were all prepared for review by the author and reflect a personal bias toward using relatively short review segments: paragraphs, at the largest; where lists are present, list elements; where large tables are presented, table cells; and individual graphic images. Section headings, bibliographic entries, and titles are usually excluded from review segments. Data collected on review segments consist of the text of the review segment and a word count.

Table III Words in Review Segments

Review Segment Size (word count)				
Characteristic	***All Types***	***Type 2***	***Type 3***	***Type 4***
Mean	24.89	35.60	21.09	73.86
Median	14	8	14	65
Standard Deviation	28.19	43.33	20.41	55.23
Sample Variance	794	1877	417	3051
Kurtosis	17	1.5	5.6	4.7
Skewness	3.2	1.4	2.1	1.7
Range	306	165	158	306
Minimum	2	3	2	2
Maximum	308	168	160	308
Count (Number of segments)	1822	48	1656	118

The sample variances of both the raw data and a logarithmic transformation are too heteroscedastistic to perform a reliable analysis of variance. A null hypothesis of no differences between the three document types cannot be rejected. Examination of the means and standard deviations points out and obvious difference between types 3 and 4. This conclusion is supported by the nature of the genres represented: type 4 documents are drafts of conventional papers dominated with paragraph-long segments; and type 3 documents are dominated by meeting minutes composed of short segments such as action items and list bullets.

Comments

Each review segment attracts a set of comments, usually an empty set. The set may include not only comments on the review segment, but also comments on the other comments on the review segment. The comments are entirely free form, either text or HTML, and may include emphasis, paragraphing and even images.

Data collected on comments includes: the text of the comment, a word count, the name of the commentator, the commentator's e-mail address, the time and date, and the qualitative coding of the comment, both Bales codes (see §5.1.4.1) and Meyers codes (see §5.1.4.2). Due to the unrestricted length of comments, the unit of analysis for coding purposes must usually be a fragment of the message. The Bales codes were assigned to comments by dividing multi-sentence comments into written equivalents of speech acts. These fragments, as noted by Henri cannot be rigidly determined, but must be parsed out based on the analytic objectives[1]. This same conclusion is seen in Meyers, *et.al*., where the units were "complete thoughts," rather than words or turns[2]. Occasionally, there are additional, usually social, meanings that can be read into the commentary. For example, the wording of a comment may contain aggressive or supportive intent.

Table IV Words in Comments

Comment Size (word count)					
Characteristic	***All Types***	***Type 2***	***Type 3***	***Type 4***	***General***
Mean	31.83	34.80	21.49	54.46	30.73
Median	19	22.5	12.5	43	12.5
Standard Deviation	37.61	36.98	26.79	48.01	36.77
Sample Variance	1414	1367	717	2305	1352
Kurtosis	14.6	1.2	40.3	7.9	0.8
Skewness	3.1	1.5	5.1	2.2	1.5
Range	289	122	256	288	124
Minimum	1	3	1	2	1
Maximum	290	125	257	290	125
Number of Comments	233	20	148	65	40

The sample variances are too heteroscedastistic to employ an F test, so a logarithmic transformation was attempted. The sample variances found in the transform were 1.18, 1.09, and 1.02. Using the transformed data a value of 22.1 was found for F. The value of $F_{2,232(.001)}$ is 6.9, so a null hypothesis of no differences between the three document types can be rejected at the .001 level. Comments made on type 4 documents are much longer on average than comments made on type 3 documents such as meeting minutes.

In analysis of the DocReview commentary, it was discovered that the DocReviews of meeting minutes constituted a subset of commentary that demonstrated very random annotative behavior (see Figure VII). When the comments for meeting minutes are removed from the Type 3 comments, the sample variances are too heteroscedastistic to

employ an F test, so a logarithmic transformation was attempted. The sample variances found in the transform were 1.18, 1.26, and 1.01. Using the transformed data a value of 2.65 was found for F. The value of $F_{2,104(.10)}$ is 2.36, so a null hypothesis of no differences between those three document types can be rejected at the .10 level.

5.1.4 Qualitative Coding Systems

Analysis of the content of the annotations must start with the selection or invention of a qualitative classification system. Many investigators have seen the wisdom of creating coding systems that are fitted closely to their problem. I chose to use existing systems, thus providing a possibility of drawing comparisons. I chose two systems, one for gauging the social functioning of research team, and the other to show how commentary became argumentation in the review of the document.

Any classification scheme must serve to differentiate between members of a group of cases. In our study the cases are DocReviews, an object that consists of a document that is partitioned into "review segments", and a set of comments made on each segment. The number of comments may be zero or more, and is usually zero. In uncommented segments, the question of implied agreement must be raised. One may be tempted to assume, since there is no limitation on reflection, that the reviewers agree with the review segment. Implied assent is very dangerous because it enables power mechanisms. No comment just means that the reviewers chose not to add to the dialog[3].

So how can we differentiate between the DocReviews? Certainly there are descriptive statistics such as size of the base document, the number of review segments, the number of comments, when the comments were made with respect to opening the review process, the size of the comments, and who made comments. These data were maintained in the log files, which are features of DocReview.

Beyond these physical statistics are the study of the character of the social interactions of the review team, the interpersonal process analysis (IPA); and the study of the efficacy of the review process, how the review contributed to the refinement of the knowledge represented by the document. Both the IPA and studies of efficacy can be conducted only by analysis of the content of the annotations. Measurement of the value of the comments to the collaboration is quite impossible in most cases, but a qualitative categorization of comments can be done by at least two classification schemes: an observational scheme and a scheme based on how the comments would fit into a formal argument. We must then code the DocReview multilogues twice, once for the social dimension of process-orientation and again for the knowledge content dimension of task-orientation.

To analyze the interpersonal process analysis of behavior in DocReviews, I classified the annotations using the Bales' codes[4], a well developed and respected tool. Analysis of how comments within a DocReview contributed to the knowledge-building content of the document will be conducted using a coding system based on the function of the comment from a task-oriented viewpoint, rather than from a social viewpoint as in IPA. The task-oriented functions are defined as the character of the comment (or comment fragment) in a formal argumentation framework. Meyers, Seibold and Brashers developed this coding system that was based on, and extended from, their previous work[5].

Classification schemes need to satisfy three conditions[6]:

- There are consistent, unique classificatory principles in operation.
- The categories are mutually exclusive.

- The system is complete [exhaustive[7]].

The coding schemes I use vary in compliance with these desiderata. Bales codes are not complete; there is no place for nonsense or muttering. The Bales codes are not mutually exclusive, they instead are derived from four fairly distinct major categories that are each divided into three quasi-ordinal codes that have very fuzzy boundaries, e.g. what is the difference between giving information and giving an opinion? Bales attempts to close the ambiguities in the codes by a very thorough explanation of each[8], but overlaps and gaps exist.

Meyers' scheme provides a less complete guide to coding[9] (appropriate for a research article as opposed to Bales' book). Both coding schemes are well described and coders can become facile with them in a reasonable time period. In reference to mutual exclusivity, a continuous system like Bales' IPA, must have fuzzy boundaries. Meyers' system is not a continuous system so is immune from this argument.

Meyers' scheme neatly solves the completeness problem with the introduction of the category "non-arguable." Fortunately, this category can contain no contextual knowledge, so it can safely be excluded from our analyses. Bales asserts that his categories are made complete and continuous by being concerned with the interaction content rather than the topical content and by eliminating any requirement for the observer "to make judgments of logical relevance, validity, rigor, etc."[10].

Correct assignment of codes could perhaps be tested by comparing actual results from dialog in the source research and the coding of the same material by the author. In short,

such testing would require studying intercoder reliability between the teams of Bales and Meyers and the team (myself) that would code the annotations. Bales offers six pages of coded dialog[11]. Meyers et. al. offers some short examples. Both papers do offer good definitions of the categories. The categories are based on dialog quite familiar to any literate individual. A larger issue is the absence of gestural side-channel communication (head nodding, eye-rolling) in DocReview. As face-to-face dialog would present frequent "speech acts" that are gestures, facial expressions, or voice tones, there will be a loss of that dialog in the coding of DocReview annotations. This loss may account for some of the significantly lower "social-emotive" codes in the DocReview annotations.

I can only compare DocReviews to DocReviews since there was no attempt to set up a control review method by other means. In the DocReview study, all DocReviews use the WWW and are thus device independent. Usually, the participants within a given set of DocReviews are homogeneous, though between sets, they may vary in number. The same task is always performed: review of a document, though the nature of the documents may change (meeting minutes, position papers). Almost all users are invited, since most DocReviews are on intranet sites. Other than the exceptions noted, most dependent variables are identical. Most studies that apply IPA compare computer-mediated communication with face-to-face communication. In a meta-analysis of studies of computer-mediated collaboration, McGrath and Berdahl[12] make several cautionary points based on differences between face-to-face interaction and computer-mediated interaction: studies often use different computer systems; different kinds of participants are used; different types of tasks are performed; and there are different patterns of dependent variables.

5.1.4.1 Interaction Process Codes

These codes are intended to assign speech acts, including backchannel communication, to categories that are based on social processes rather than substantive content. Since we are social animals, the nature of our dialog will to a great extent determine both how we respond emotionally to our collaborative environment and how effective that environment is in attracting productive participation.

The Bales Code

Commentary of hyperdocuments through DocReview can be evaluated by use of categorization, volume and quality. DocReview comments can be categorized by using Bales codes[13]. Depending on the issue domain, these codes can be used to order value between categories. For instance, detection of errors in spelling or grammar is a low value contribution in studies of social behavior, but a high value contribution in the development of a manifesto or epic.

Table V Bales Code for Acts in Social Interaction

Main Categories	Frequency	Types of Acts	Frequency
Positive reactions	25.9%	Shows solidarity	3.4%
		Shows tension release	6.0%
		Shows Agreement	16.5%
Problem-solving attempts	56.0%	Gives suggestion	8.0%
		Gives opinion	30.1%
		Gives information	17.9%
Questions	7.0%	Asks for information	3.5%
		Asks for opinion	2.4%
		Asks for suggestion	1.1%
Negative reactions	11.2%	Shows disagreement	7.8%
		Shows tension	2.7%
		Shows antagonism	0.7%

--- after Bales, 1955

Commentary that expresses support or disagreement is not valueless, for such commentary does influence the behavior of the author and other contributors. So most commentary is of some value, even if it is merely reinforcing the recognition of a team effort. Sadly there are comments of negative worth that occasionally emerge, such as personal attacks or senseless graffiti.

Gay et.al. and classroom discussion forums

Geri Gay and others studied the character of student contributions by computer-mediated

communication in university classes[14]. The discussion forums were conducted in CoNote, a WWW-based annotation program functionally similar to DocReview. Gay's study included questionnaires and observer data as well as a repository of documents and comments thereon. Gay's codes, like Bales' codes, are not based on the relationship of the annotation to the collaboration task, but on the character of interpersonal activity. Content of the annotations was organized into three categories: technical comments, affiliative comments and advice. Presumably, a single comment could contain all categories, but not multiple occurrences of a category. 197 comments produced percentages of 50.3 technical, 45.2 affiliative, and 68.5 advice. These percentages were obtained in an environment dominated by students who came into frequent contact, thus by age and group structure more inclined to engage in affiliative commentary than professional groups might be.

These codes are equivalent to portions of the twelve category Bales Codes for Interpersonal Activity. The *affiliative* comments, which presumably could be positive or negative, would fall into one of six categories: Shows Solidarity, Shows Tension Release, Agrees, Disagrees, Shows Tension or Shows Antagonism. The *technical* comments would fall into the neutral task-oriented area: Gives Opinion, Gives Orientation, Asks for Orientation, Asks for Opinion. The *advice* category corresponds to the extreme range of the task-oriented area: Gives Suggestion and Asks for Suggestion.

5.1.4.2 Argumentation Based Codes

If research is analogous to argumentation, as Eisenhart and Borko suggest[15], then a coding system that is based on the argumentation process would seem to be a more effective alternative for characterizing task-oriented activity than the more process-oriented Bales IPA coding. The value of a comment fragment (coding unit) to the collaboration is more closely related to task than process. Perhaps we can assign a value

to a specific type, or if the coder is familiar with the document, we can actually assign an interval measure for value. Three coding systems have been considered: informal argumentation codes, structurational argument codes, and an observational categorization.

Informal Argumentation

In *An Introduction to Reasoning,* Toulmin, Rieke and Janik develop a dialog classification based on argumentation[16]. Their system is proposed to be the basis for development of a tool (The Landscape of Reason) to organize dialog for the Research Web. Argumentation is broadly defined in this work, having a place in any "rational enterprise." As the authors put it, "... scientific arguments are sound only to the extent that they can serve the deeper goal of improving our scientific understanding." Every coding unit of a comment can be assigned a type based on this classification. The value of the comment in terms of value to the collaboration can be established through a surrogate, the value of the comment in the argument. There are six elements in argumentation: claims, grounds, warrants, backing, modal qualifiers, and rebuttals.

- Claims are assertions put forward publicly for general acceptance. In DocReview terms, every review segment is a claim. The claim is that the review segment, whatever its nature within the document, presents and argues its proposition well, and conforms to accepted standards. For example, the role of a review segment within the document may be that of, say, a rebuttal. Comments directed toward the review segment can, in a recursive way, present grounds, warrants, backing, modal qualifiers and rebuttals to the review segment's basic claim *as a rebuttal*.

- Grounds are facts that support a claim. Comments may be directed toward the grounds given, in our example the grounds supporting a rebuttal.

- Warrants are ways of describing how one can validly draw a conclusion from the grounds offered. This is the argument in argumentation.

- Backing makes explicit the experience relied on to establish the trustworthiness of the warrants. In scholarly documents, citations of literature are the principal means of supplying backing.

- Modal qualifiers are statements that show what kind and degree of reliance is to be placed on the conclusions.

- Rebuttals are statements showing exceptional circumstances where the conclusions might be undermined.

Structurational Argument Codes

In research on decision-making discussions in a face-to-face environment, a set of seventeen categories describing statements in terms of their place in argumentation was developed and used by a team that studied 45 discussions. This research had its roots in research by Toulmin (in 1958) and two other research teams in 1969 and 1980[17]. I can find no subsequent application of this coding scheme in the literature. Coding is extremely difficult, as meanings can shift with context. The coder must be thoroughly immersed in the argument, not just the words, but also the intent of the words.

In Meyers *et.al*. discussions were analyzed with 8,408 codes produced, having the distribution given in the following table[18]. This dissertation found 425 codes in the DocReview annotations.

Table VI Structurational Argumentation Codes

ARGUABLES (67.4%)	Potential Arguables	***Assertions***	Statements of fact or opinion.
		Propositions	Statements that call for support, action or conference on an argument-related statement.
	Reason-using Arguables	***Elaborations***	Statements that support other statements by providing evidence, reasons, or other support.
		Responses	Statements that defend arguables met with disagreement.
	Reason-giving Arguables	***Amplifications***	Statements that explain or expound upon other statements in order to establish the relevance of the argument through inference
		Justifications	Statements that offer validity of previous or upcoming statements by citing a rule of logic (Provide a standard whereby arguments are weighed).
REINFORCERS (13.6%)		***Agreement***	Statements that express agreement with another statement.
		Agreement +	Statements that express agreement with another statement and then go on to state an arguable, promptor, delimitor, or non-arguable.
PROMPTORS (2.3%)		***Objection***	Statements that deny the truth or accuracy of any arguable.
		Objection +	Statements that deny the truth or accuracy of any arguable and then go on to state another arguable, promptor, delimitor or nonarguable.
		Challenge	Statements that offer problems or questions that must be solved if agreement is to be secured on an arguable.

Table VI (continued)

DELIMITORS (2.1%)	***Frames***	Statements that provide a context for and/or qualify arguables.
	Forestall/ Secure	Statements that attempt to forestall refutation by securing common ground.
	Forestall/ Remove	Statements that attempt to forestall refutation by removing possible objections.
NONARGUABLES (14.5%)	***Process***	Non-argument related statements that orient the group to its task or specify the process the group should follow.
	Unrelated	Statements unrelated to the group's argument or process (tangents, side issues, self-talk, etc.)
	Incomplete	Statements that do not provide a cogent or interpretable idea (due to interruption, stopping to think in midstream, but are completed as a cogent idea elsewhere in the transcript.

after Table 1 Meyers, Seibold and Brashers[19]

While Meyers et.al. conclude that the structurational argumentation codes reflect *both* process-orientation and task-orientation (or system and structure, as they put it); the coding scheme clearly supports task-orientation much better than the Bales IPA. In terms of support to a collaborative task, some categories have more value than others.

These argument codes provide places for every element in the Toulmin informal argumentation scheme. The nonarguables Process and Unrelated are very convenient "bins" for trivial or procedural content. One of the seventeen codes is extremely unlikely to be used: the nonarguable *Incomplete*. The argument codes were developed to analyze transcripts of face-to-face interactions, an environment where interruptions are frequent. It is difficult to imagine how an asynchronous contribution could be interrupted; if the

writer is interrupted at the terminal, then the task can be resumed when the interruption terminates.

The Meyers, et.al. study used transcripts of actual face-to-face multilogue, with recourse to videotape only when the expression needed clarification[20]. Interruption and incomplete expressions were frequent, as in normal conversation. The computer-mediated environment of DocReview will make interruption unlikely and incomplete thought rare. I expect the distribution of message fragments in DocReviews to be quite different from conversational multilogues. As McGrath and Berdahl cautioned, these differences may be due to many different factors[21]; nevertheless, if the differences are great, the argument in favor of computer-mediated communication as a more reflective medium gains support.

An Observational Categorization

The author's five years of experience in the use of DocReview has led to a potential coding system based on observation and sorting. Interpretation and characterization of the codes are based not only the original context of the commentary, but on assumptions of what character the comments would take in a fully implemented Research Web.

This scheme categorizes several nominal classes of comments seen in DocReviews. It has the advantage of being completely specific to DocReviews; that is it is not time restricted, and is asynchronous, document-centric. Most DocReview review segments, especially paragraphs, will contain an assertion, a conclusion and give evidence showing how the conclusion follows from the assertion. In addition to this logical imperative (substantial) there is also the requirement to conform to appropriate standards of

scholarship and presentation (formal). In the Research Web environment, the documents are also subject to both the criticism process and an editing process.

Table VII

Observational Categorization of DocReview Annotations

	Substantial	***Formal***
Editing Process	• Supplies references and citations. • Supplies new information or examples. • Suggests deleting information.	• Corrects spelling or grammar. • Questions document layout. • Questions sentence or paragraph structure.
Criticism Process	• Questions validity of statements. • Gives opinions or suggestions. • Supports or rejects substance with grounding.	• Supports a comment w/o supplying new information. • Disagrees with a comment w/o supplying new information. • Questions or discusses the philosophical bases of the document.

5.1.5 Qualitative Coding Reliability

In the analysis of the data, the distribution of codes in the DocReview commentary is compared to the distribution of codes in the studies that defined the codes. In comparing the distributions, there is the necessary assumption that all coding would be consistent and correct. Bales points out three sources of variation between coders: unitizing, the correct parsing of dialog into units of analysis; categorizing, correct assignment of codes; and attributing, the source and target of the dialog[22]. There is, in the DocReview analysis, no question of the source and target. Because this dissertation was not well funded, the author did all coding, so the skill and consistency of coding was not established by comparing the coding of dialog by independent coders.

Unitizing is a significant source of variability. The variability in unitization is induced by uncertainty in interpretation. Some methods of unitizing are less susceptible to variability than others. Time-based unitization, segments of elapsed real time, are not subject to interpretation[23]. Turn taking in speech dialog is more variable due to complications that arise in parsing of monologues; annotations in DocReview are essentially monologues. Parsing face-to-face dialog into speech acts (Bales) is yet more variable because there is a need for insertion of implied speech acts and gestural acts. Even more variable is the event-based coding that was used in the argumentation coding (Meyers). Nyerges *et.al*. chose time-based coding over event-based coding because event-based coding required at least two coding passes[24].

In the Bales coding, DocReview annotations were parsed during coding into approximations of "speech acts" by dividing the annotation into phrases, sentences or a set of contiguous sentences that dealt with a single topic. Not infrequently when the coder understands both the review segment and an annotation well, implied codes emerge. One comment usually contained a few codes (mean = 2.6) sometimes as many as a dozen. This parsing is assumed to be equivalent to the turn taking of face-to-face dialog.

In the argumentation coding, the unitizing protocol used in Meyers et.al. could not be employed since their unitizing was done by two judges concurrently. As Meyers used transcripts of dialog, so I used written dialog. The unitizing rule that Meyers et.al. used was: "any statement that functioned as a complete thought or change of thought." The Meyers team coded dialog that was parsed into turns, while DocReview comments are relatively long monologues. Rather than parsing the monologue into speech acts I parsed it into argument units that might include several sentences. Such units fit well into the

Meyers categories. One comment usually contained one to a few codes (mean = 1.4) sometimes as many as eight.

Coding and unitization of DocReview annotation requires the coder to place the annotation into the context of the review segment being annotated. This contextualization is done by mentally converting the annotation unit and review segment into a narrative equivalent. Unfortunately, returning to the exact same mind set is difficult for either independent judges or for the same coder repeating the coding at a later time.

5.1.5.1 Coding Reliability Tests

In order to test the reliability of the coding, it was decided to take a 12.5% random sample of all review segments that received comments. The author, who was the original coder of the entire set of comments, then recoded this sample. There was no recall of the original coding.

Four sets of codes were tested for reliability: the Bales codes (twelve categories), the Bales categories (four sets of three codes each), the structurational argumentation codes (seventeen categories), and the five structurational argumentation categories derived from the seventeen codes.

5.1.5.2 Data Conditioning

The parsing of DocReview annotations into coding units (unitizing) proved to be uncomfortably variable. It seemed that the degree of engagement by the coder was the principal source of variability. When coding the annotations, the context had to be set by

reading and understanding the review segment and then interpreting the annotation in context. When the coder is well engaged, the dialog shows more nuances (codable) than a perfunctory reading would provide. Of course coder drift and fatigue contributed to the variability too. If the coder is heavily engaged in reading between the lines, and sees and records an implied code that another does not, then there will be a difference in the number of codes. The two code strings may not align: for instance coder 1 codes "acbbbca" and coder 2, missing the implied code, codes "cbbbca".

Aligning codes at the beginning gives:

acbbbca
cbbbca

This results in 2 agreements, 4 disagreements, and one not matched.

If on the other hand we align like this:

acbbbc
a
cbbbca

Then we have a probably more accurate analysis of six agreements and one not matched.

If such realignment is allowed it is subject to much abuse, so I allow only a shift of the entire shorter code string within the limits of the longer code string. If the code strings are of equal length, then no shifting is allowed. Any unmatched codes resulting from unequal code string lengths are removed. Both Bales and the structurational argumentation codes were conditioned this way, and the resulting conditioned data was converted to the aggregated categorical data (the four Bales categories and the five structurational argumentation categories).

5.1.5.3 Analysis

Intercoder or recoder reliability can be measured by several methods. Cohen[25] and Landis & Koch[26], in their examples, use nominal categories that are clear, complete and

mutually exclusive. On the other hand Perreault and Leigh use more qualitative (though unstated) codes[27]. On this basis, plus favorable arguments from the Meyers et.al. paper, I am inclined to use the Perreault and Leigh measure. Since Cohen's kappa is so widely used, I include it for comparison purposes.

The conditioned data were placed in contingency tables comparing the two coding sessions. From the contingency tables, Cohen's kappa and Perreault and Leigh's Index of Reliability were calculated for the four sets of data.

Bales codes

From the initial set of 99 Bales codes, there were 82 codes remaining in the conditioned data. Each code could assume one of twelve values. Comparing the two sets showed 54 pairs in agreement, 28 pairs in disagreement and 17 unmatched codes. Cohen's kappa[28] for the Bales codes is 0.538, showing only moderate agreement between the two coding sessions[29]. The Index of Reliability[30] is 0.792 with a 95% confidence level of +/- 0.088. This mediocre result, in conjunction with some very low counts of several codes, provided the argument to use only the four Bales categories in the analysis of DocReview annotations.

Bales categories

In analyzing the four Bales categories, each code could assume one of four values. Comparing the two sets showed 80 pairs in agreement, 2 pairs in disagreement and 17 unmatched codes. For the Bales categories, Cohen's kappa is 0.878, showing almost perfect agreement between the two coding sessions. The Index of Reliability is 0.984 with a 95% confidence level of +/- 0.027.

Structurational argumentation codes

From the initial set of 70 structurational argumentation codes, there were 48 codes remaining in the conditioned data. Each code could assume one of seventeen values. Comparing the two sets showed 21 pairs in agreement, 27 pairs in disagreement and 22 unmatched codes. Cohen's kappa for these codes is 0.402, showing only fair agreement between the two coding sessions. The Index of Reliability is 0.668 with a 95% confidence level of +/- 0.133. As with the Bales codes, there were a large number of codes with low to zero counts.

Structurational argumentation categories

In analyzing the five structurational argumentation categories, each code could assume one of five values. Comparing the two sets showed 28 pairs in agreement, 20 pairs in disagreement and 22 unmatched codes. Cohen's kappa is 0.383, showing only fair agreement between the two coding sessions. The Index of Reliability is 0.673 with a 95% confidence level of +/- 0.133.

5.1.5.4 Conclusions

It was presumed that the Bales codes would measure the social aspect of DocReview "dialog." The reliability of the coding was acceptable, especially for the four Bales categories. It must be noted however that the strong concentration of the codes into the positive task-oriented category results in reliability that is perhaps misleading.

The structurational argumentation codes were too numerous and difficult to code to produce acceptable reliability. Applying argumentation codes to analysis of DocReview annotations will require the use at least pairs of coders working together (as Meyers et.al. did). The unitization problem was extremely serious, producing almost a one third rate

of no matching codes. The combination of arbitrarily long review segments and arbitrarily long annotations will demand a very clever unitization scheme to produce any hope of consistent coding.

5.1.6 Analytical Results

The proposition designations below (e.g. A2) refer to the research questions discussed in §5.1.1. Three techniques were used to test the propositions: Chi-squared, regression analysis, and case studies.

Four of the propositions use the Chi squared test comparing the counts of DocReview codes versus the coding distributions in the original Bales and Meyers studies. In order to normalize the sample sizes a pseudo-sample of the Bales or Meyers codes was drawn with the same distribution as in the original studies but with a size equal to the DocReview sample.

Four of the propositions were tested using single variable regression analysis. In all these cases the independent variable (X) was the word count of the base document or a review segment of the base document. In some cases the dependant variable (Y) was confounded with the independent variable. This confounding was due to the definition of effectivity as the ratio of commentary to the size of the document (effectivity = Y/X). The shape of the best fitting regression line was found to be logarithmic.

One of the propositions was a case study comparing DocReview to three other web-based annotation programs. The comparison was made on the basis of a universe of features found in all the programs.

5.1.6.1 Proposition A1. The social character of comments in DocReview differs from comments in face-to-face dialog.

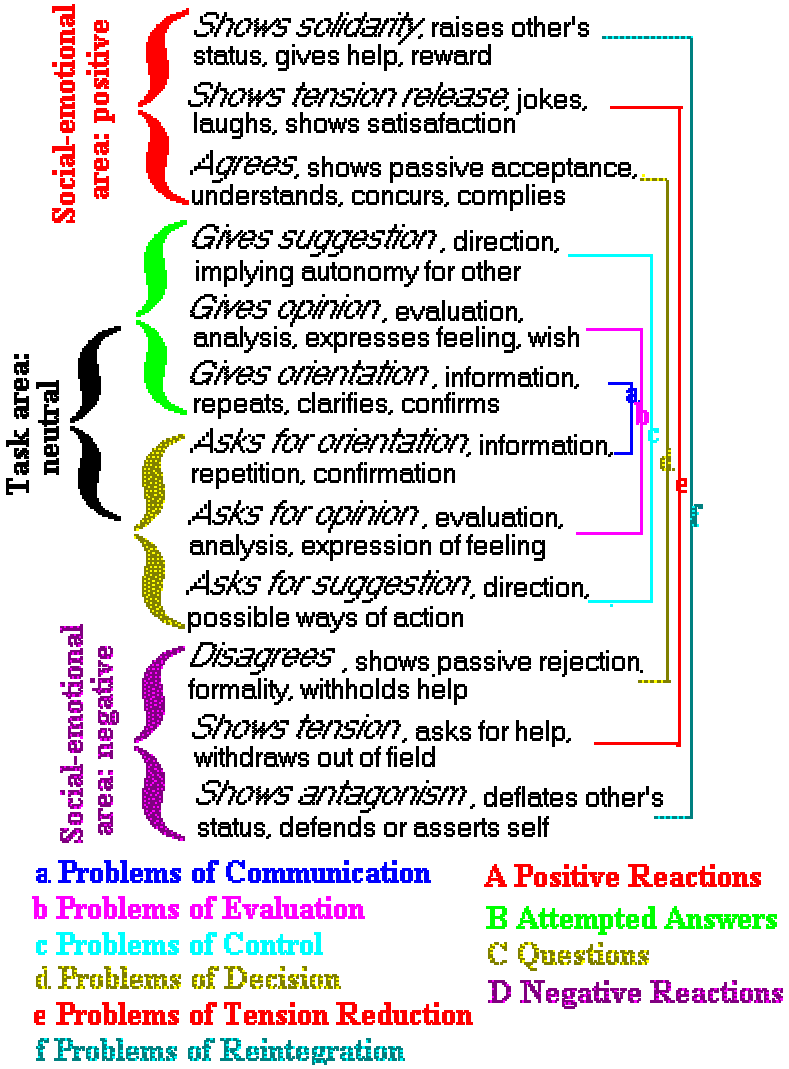
One of the most important questions arising from the use of DocReview is how the nature of dialog in DocReview is different from face-to-face dialog. Fortunately we have from Bales' work a distribution of codes assembled from thousands of face-to-face speech acts. If one makes the assumption that DocReview annotation is equivalent to one side of a face-to-face dialog, and further assume that in face-to-face dialog the two participants each produce an identical distribution of coded speech acts, then we can make a valid comparison. The assumption of equivalence is strained by the odd nature of this communication: essentially the document is the source of a series of propositions. The annotation is a set of responses to the proposition presented in the review segment by the readers. This set of responses is also complicated by the not infrequent presence of commentary on other annotations.

Operationalization:

Assigning Bales codes categories to all annotations operationalizes the social character of the comments. The Bales Interaction Process Analysis categorizes all speech acts, including gestures, into twelve codes. The differences between some of the Bales codes are very slight. These fine nuances result in a high variability between coders or between coding sessions by the same person. In order to reduce the intercoder variability it was decided to use Bales' broader classification: categories. Bales grouped the twelve codes into four categories that are generic and form a good basis of comparison. These categories are: positive reactions, problem-solving attempts, questions, and negative reactions. Problem-solving Attempts and Questions are further generalized into a

supercategory of the task area, while Positive and Negative Reactions are generalized into the social-emotive area.

Table VIII Bales Interaction Process Analysis Codes



from (Bales 1950)

Data conditioning:

None.

Data Analysis:

The counts of codes of the entire set of DocReview annotations by Bales category demonstrates that DocReview annotations show a much higher degree of task-related dialog and a much lower degree of social-emotive dialog than is seen in face-to-face dialog. The comparisons (DocReview/face-to-face) are: for Negative Reactions -- 0.1%/11.2%; for Questions -- 7.3%/7.0%; for Problem-Solving Attempts -- 85.5%/56.0%; and for Positive Reactions -- 7.0%/25.9%.

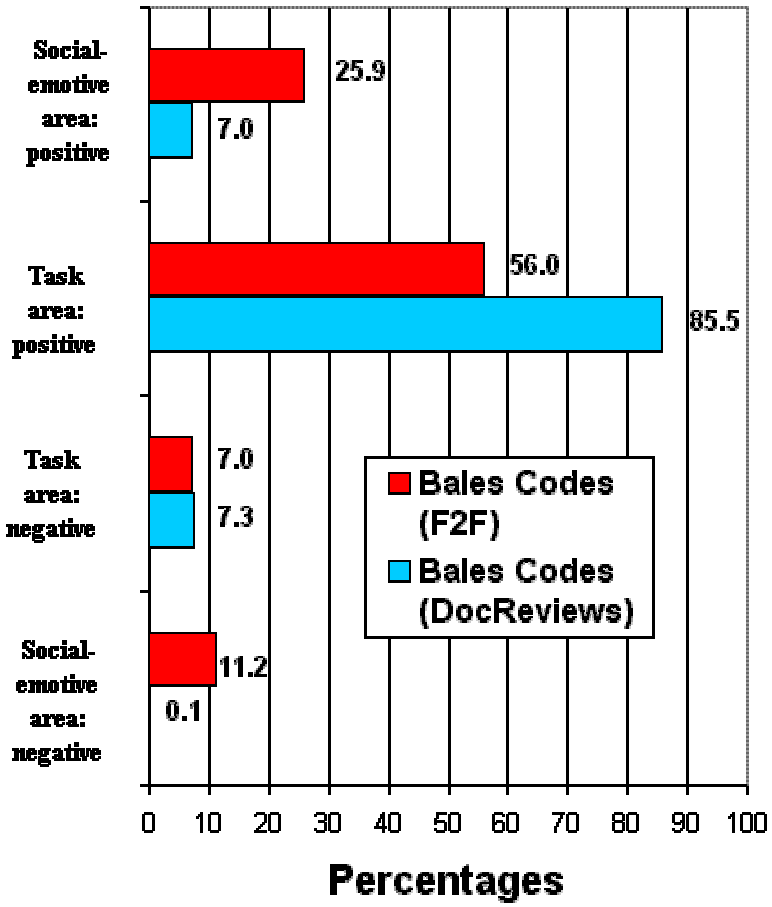


Figure I Distribution of Bales Codes

We find that the null hypothesis that there will be no difference between face-to-face and DocReview dialog when Bales coded can be rejected. With three degrees of freedom, Chi squared = 213.2. This result is significant at <0.000001.

Discussion of Findings:

The very low percentages of DocReview annotation in the social-emotive area (positive and negative reactions) may show the effect of moderation in dialog induced by the reflection afforded by DocReview as opposed to the more spontaneous nature of face-to-face dialog. The similarly low, though less extreme, percentages in the positive reactions category may show that there is less need felt for social reinforcement than in face-to-face dialog. Though DocReview annotations show less positive reinforcement, the reinforcement is there, it is simply less effusive. Questions (task area: negative) show an almost identical distribution. Problem-solving attempts (task area: positive) are much higher in DocReview annotation than in face-to-face dialog. This disparity may be the result of the ability of the reader to reflect much longer than is possible in face-to-face dialog. I suggest that this is the most important finding, demonstrating the value of DocReview in problem solving.

Interpretation of Findings:

The conclusions must be tempered with the realization that there are no gestural acts in the DocReviews and their annotations. While Bales does not record the percentages of gestural acts captured in his research, in his description of the codes gestures such as winks, nods, frowns, and even blushing appear. From Bales' description of the codes one can clearly see that most gestural acts are in the social-emotive categories. If an arbitrary portion of the Bales social-emotive codes (comprising 37.1 % of the total face-to-face acts) was assumed to be gestural, then in the annotation coding the missing

percentage would need to be reassigned from the task oriented categories. This reassignment would cause the comparison between positive task-orientation to be somewhat less marked, and the comparison between negative task-orientation would shift from being almost equal to somewhat less negative than in face-to-face dialog.

5.1.6.2 Proposition A2: The substantive character of comments in DocReview differs from comments in face-to-face dialog.

The substantive nature of comments in DocReview is measured by determining the intent of the comment, or a portion of the comment. Intent is defined in this analysis as what place the comment would take in argumentation.

As in the analysis of social character of the comments above in Proposition A1, we have to assume that the dialog is quite one-sided, with the document providing propositions and the readers arguing with that proposition. Clearly there can be no negotiation of meaning and the document can make no rebuttals. In terms of argumentation, then we can have but one round of argumentation, but with several people participating.

Operationalization:

Assigning Meyers structurational argumentation code categories to each comment operationalizes the substantive character of the comments.

Data conditioning:

The raw data percentage comparisons (DocReview/face-to-face) are: for non-arguables -- 22.6%/14.5%; for delimitors -- 8%/2.1%; for promptors -- 23.1%/2.3%; for reinforcers -- 10.3%/13.6%; and for arguable -- 36%/67.4%.

Argumentation codes in the non-arguable category in the dialog were excised. In the raw data, DocReview annotations were 22.6% non-arguable, compared to 14.5% in the Meyers study. The difference in non-arguables is attributed to the assignment of annotations frequently complaining about grammar and spelling to that category. Arguably such commentary does not contribute to productive argumentation, and furthermore such corrections are seldom made in face-to-face dialog.

Codes in the arguable class were also excised. Difficulties in adjusting for the asymmetrical nature of DocReview argumentation are simply insurmountable. In the one turn dialog, responses to propositions (the base document's review segment) are much more prevalent than responses to annotations. Responses to annotation usually requires re-reading the comments; busy participants are not likely to return to review comments, even if they are reminded by e-mail notification. This would not be the case in face-to-face argumentation.

The data conditioning leaves us with three categories of codes: Reinforcers, Promptors and Delimitors. Unfortunately the excision of troublesome categories reduces our number of data points by 58% to 176. Since the central action of argumentation is carried out in these categories, I feel that they are an adequate basis for comparison.

Data Analysis:

The conditioned data comparisons (DocReview/face-to-face) are: for reinforcers -- 25%/75.6%; for promptors -- 55.7%/13.3%; and for delimitors -- 19.3%/11.1%.

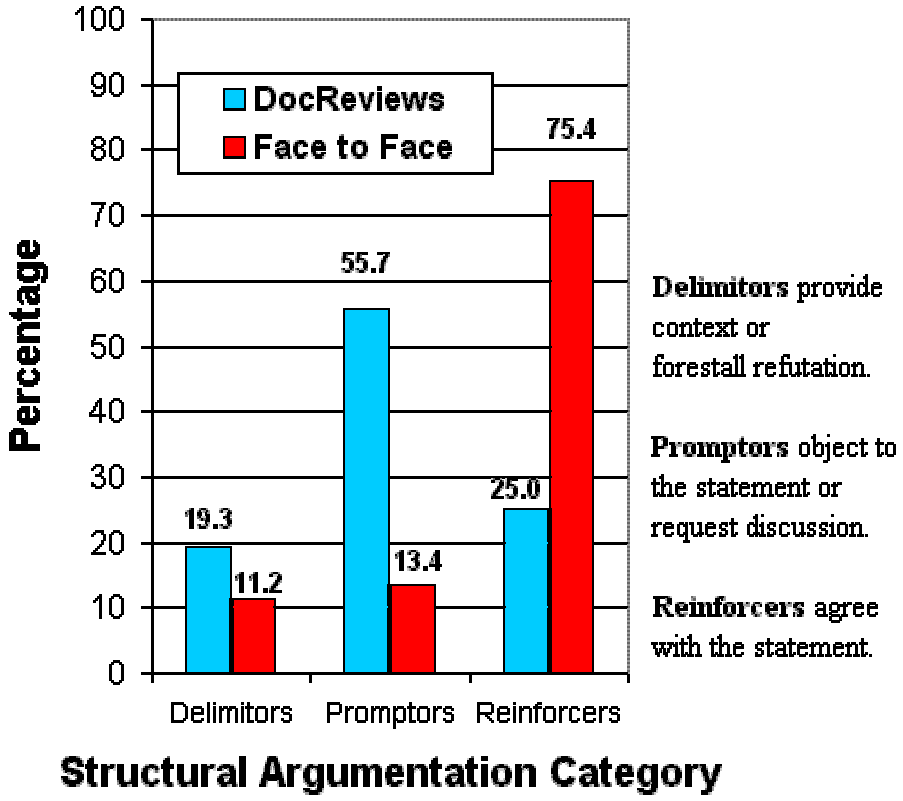


Figure II Substantive Commentary (DocReview vs. F2F)

Comparing face-to-face distributions to the distributions found in the DocReviews shows a very strong difference in both promptors and reinforcers. There are four promptors in DocReviews for each face-to-face promptor and three face-to-face reinforcers for every DocReview reinforcer.

We find that the null hypothesis that there will be no difference between face-to-face and DocReview dialog when Meyers coded can be rejected. We find that with two degrees of freedom, Chi squared = 93.3. This result is significant at <0.000001.

Discussion of Findings:

The differences of face-to-face argumentation and DocReview annotation are clear: people are much more inclined to suggest changes to the document in DocReview than in face-to-face dialog; people are much less inclined to agree with the document in DocReviews than they are in face-to-face dialog. I see this finding as suggesting that there may be some satisficing occurring as people are less inclined to annotate texts that they see as not far enough wrong to complain about. The vast difference in promptors may be explained by the nature of DocReview: documents are mounted with the intent of drawing out errors and omissions. A portion of the differences may also be explained by social mechanisms: it is much easier to praise than object; and power effects may also be seen as people are more inclined to agree with a proposition offered in a meeting (usually by a leader).

5.1.6.3 Proposition B1: Long base documents are ineffective relative to short documents.

The lives of researchers are fragmented into scores of tasks of varying importance. This produces the need to engage in multitasking, a mosaic of activity that fills the available time with periods of variable lengths. There will be short periods to review documents, provided they are of a size that will fit into the time slot. Very long documents may encourage a shallow reading, thus shallow and short commentary.

Operationalization:

Effectivity is operationalized as the ratio of the sum of comment size to the size of the base document. Size of comments and base documents are both established by software that counts the words of more than three characters. For each DocReview that attracted annotation (n = 78), the word counts for annotations to segments that attracted annotation for each DocReview were accumulated in one column and the word count for the DocReview was placed in another. The DocReview word count was plotted on the X-axis and the effectivity on the Y-Axis.

Data conditioning:

Records for DocReviews that attracted no commentary were excluded. A DocReview with segments containing graphics was excluded due to the low word count in the segments, and the heavy annotation of the segments. The same DocReview contained an anomalously long general comment.

Data Analysis:

A correlation of 0.665 on the logarithmic regression line confirms the hypothesis. With 77 degrees of freedom a value of F = 60.1 was found. As expected the slope was negative with $P = 3.27 \times 10^{-11}$. The P value of the intercept was 1.64×10^{-12}. A study of DocReviews by document type (§5.1.6.7) suggests that the logarithmic relationship is even stronger among base documents that are not meeting minutes.

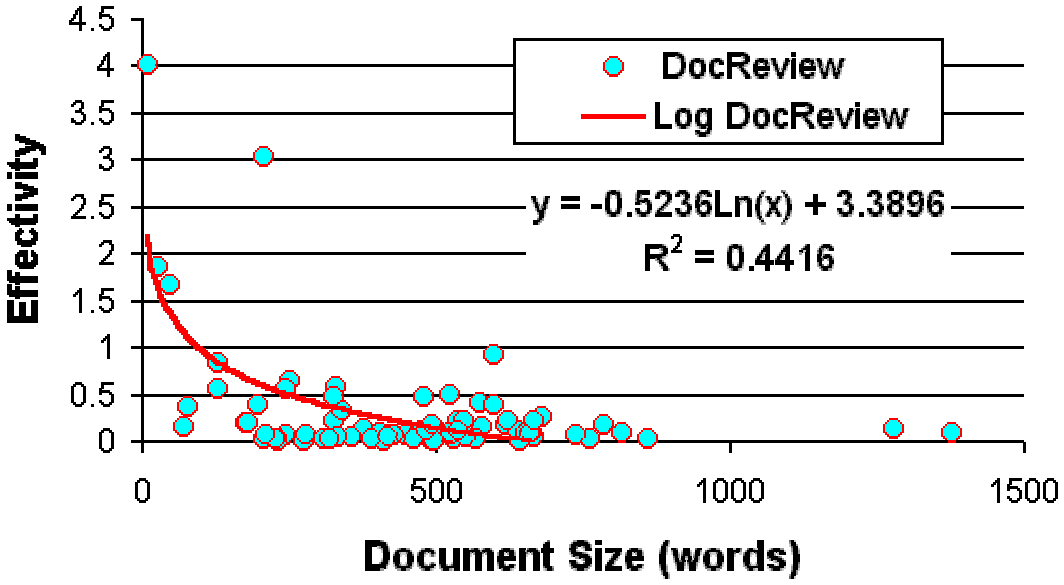


Figure III Effectivity by Document Size

Discussion of Findings:

The hypothesis is accepted. Smaller base documents produce more effective DocReviews. This leads to the conjecture that fragmenting a very long document will increase the effectivity of the review process. This conjecture could be tested, but not with the data from this study.

5.1.6.4 Proposition B2: The amount of commentary received on a review segment will be directly proportional to the segment's length.

An extremely long review segment may tax the reader's concentration, leading to a decline of effectivity. Short review segments such as list "bullets" are sharply focused and easy to grasp and critique. Due to a small denominator, the effectivity of such short

segments may be inflated. The deleterious effect of long review segments is one of the basic assumptions of the design of DocReview.

Operationalization:

Size of comments and base documents are both established by software that counts the words of more than three characters.

Data conditioning:

Segments not attracting annotation are removed. Segments that were graphic images were discarded. General comment segments were discarded.

Data Analysis:

A correlation of 0.235 on the linear regression line weakly confirms the hypothesis showing a direct relationship between segment size and received annotation. With 49 degrees of freedom a value of $F = 2.80$ was found. As expected the slope was positive with $P = 0.101$. The P value of the intercept was 0.391.

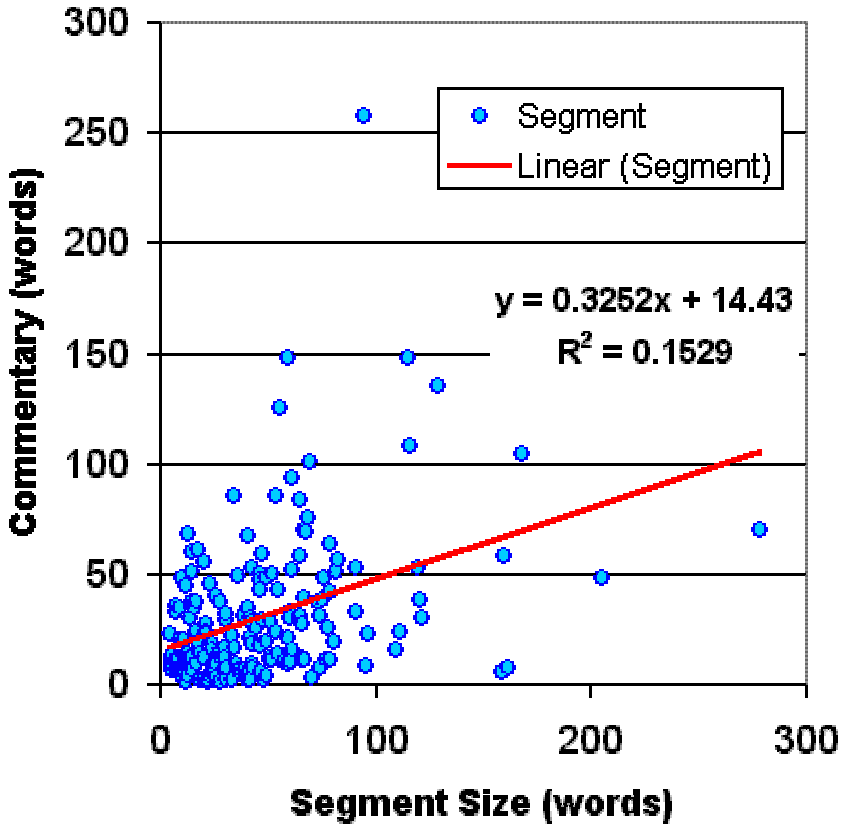


Figure IV Commentary Accumulated by Segment Size

Discussion of Findings:

The hypothesis is accepted. Commentary size is directly proportional to segment length; but while larger segments attract more commentary due to the positive slope, but they are not necessarily more effective (see §5.1.6.5) as seen by the low value (<1.0) of the slope of the regression line.

5.1.6.5 Proposition B3: The ratio of size of comments received to size of review segment (effectivity) will decline in proportion to review segment size.

Short entries in lists and cells in tables are very sharply focused and when they attract annotation, the annotations are likely to contain more information than the entry (effectivity > 1.0). The context of lists and tables are usually quite clear and contributes to their focus. When long segments such as paragraphs receive annotation, they are likely to contain less information than the segment.

Operationalization:

Size of comments and review segments are both established by software that counts the words of more than three characters.

Data conditioning:

For this analysis general comment segments were excluded, as they are not focused review segments. Segments that applied to graphic images were removed because the number of words in the graphic segment is simply the number of words in the title, and a picture is indeed often worth a thousand words. At this point outliers were examined and one more point was removed. This outlier was a document section heading that drew much commentary from the review segments within the section. Making a section heading a review segment is an error on the part of the facilitator; section headings are for ease of reading and are devoid of real content.

Data analysis:

The remaining segments that received comments were selected and two columns were produced by database query: size of the segment and the summation of the size of the commentary on the segment. This table was imported into the spreadsheet. For each segment, the size of the commentary was divided by the size of the segment to yield effectivity. A column was created for the effectivity. An XY scattergram was produced with segment size on the X-axis and effectivity on the Y-axis. A correlation of 0.451 on the logarithmic regression line confirms the hypothesis. With 184 degrees of freedom a value of F = 46.8 was found. As expected the slope was negative with P = 1.1×10^{-10}. The P value of the intercept was 1.1×10^{-18}.

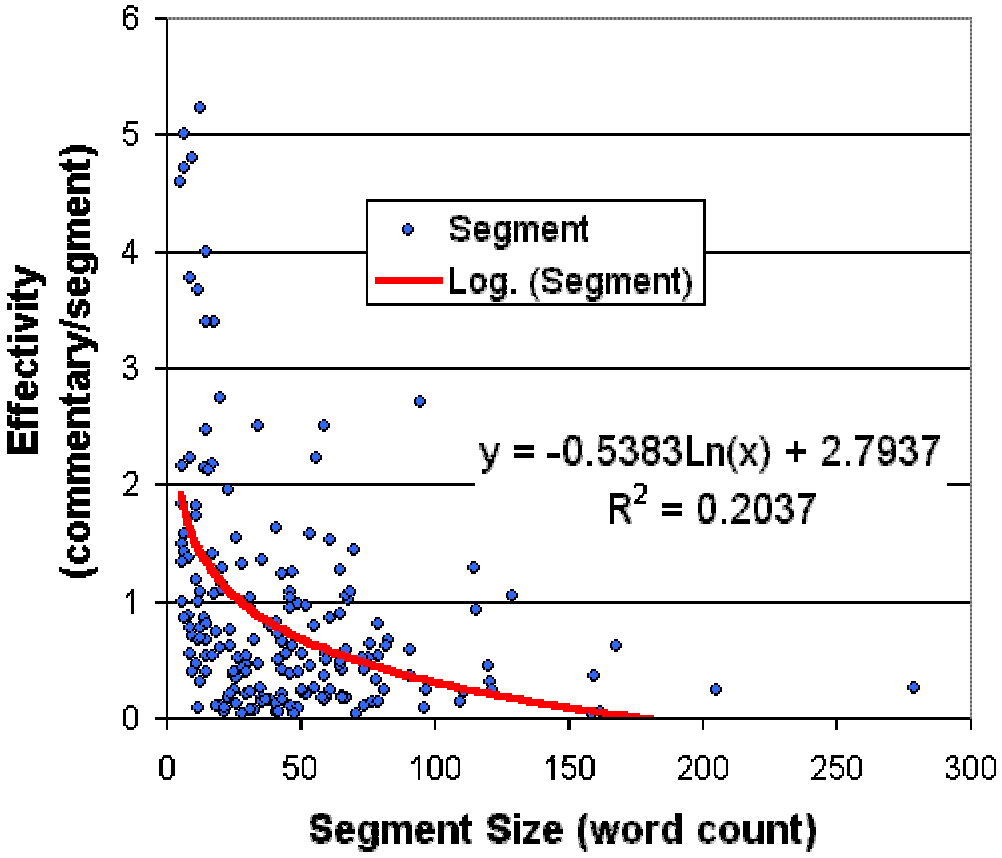


Figure V Effectivity by Segment Word Count

Discussion of Findings:

The hypothesis is accepted, with strong indications that effectivity decays logarithmically rather than linearly. This hypothesis is also supported by style guides for printed text[31, 32] and for the WWW[33]. Long paragraphs are problem-laden when reading from a screen: scrolling may be required, especially when small displays are used and when the user has the font size increased to compensate for poor eyesight. When the user has set the window to single column width, even moderate length paragraphs may need to be scrolled.

5.1.6.6 Proposition C1: Products similar to DocReview will emerge and will, by similarity, validate the design.

At least four other web-based annotation products have been put into service. One of these (Third Voice) was forced to withdraw after it was subjected to numerous lawsuits centered on copyright issues, specifically allowing anyone to copy any publicly available web page on someone else's web site for annotation.

Since DocReview's debut in 1995, three similar products have emerged: Living Documents in 1998, PageSeeder in 2000, and QuickTopic in 2001. The four products may be compared on a set of core features. The core features are: notification service, in-line commentary option, security, segmentation flexibility, comments on comments, general comments, and review all comments.

Operationalization:

The three products are compared on a set of core features.

A DocReview demo may be used at
http://students.washington.edu/~veritas/DocReview/review30.cgi?name=DrDemo.

Several Interactive Papers may be examined at http://lrsdb.ed.uiuc.edu:591/ipp/.

A Document Review may be examined at
http://www.quicktopic.com/6/D/QXx3sZA2kptQpnq9Rqwv.html.

A PageSeeder demo may be used at
http://ps.pageseeder.com/ps/ps/demos/tryit/choco/choco.pshtml.

Table IX Annotation Program Features

	DocReview	*Interactive Papers*	*QuickTopic Document Review*	*PageSeeder*
Notification Service	Yes	No	Yes	Yes
In-line Commentary	Yes, click for alternative format.	Yes, by request.	No	Yes, no other alternative format.
Security	Yes, your server.	Yes, your server.	By obscure URL.	Yes, commercial service.
Segmentation Flexibility	Yes	No	No, paragraphs and list elements only.	No, chunks only.
Comments on Comments	No, by design.	Yes, three deep.	No, by design.	Yes, unlimited.
General Comments	Yes	Yes	Yes	No
Review all comments	Yes	No	Yes	No

Discussion of Findings:

DocReview's design has been validated by the similarity of several commercial and academic products that were developed in the five years following DocReview's original release.

5.1.6.7 Proposition D1: Higher quality documents will attract more participation.

Document quality may be categorized on an ordinal scale. Degree of completion on a scale from conceptual sketches to completed canonical documents. We have categorized the documents on a five-valued quality scale (see §5.1.2).

Operationalization:

Participation is considered equivalent to effectivity and is operationalized as the ratio of the sum of comment size to the size of the base document. There were three document types represented: types 2, 3, and 4.

Data Conditioning:

DocReviews without comments were discarded. A DocReview with segments containing graphics was excluded due to the low word count in the segments, and the heavy annotation of those segments.

Data analysis:

Table X Effectivity of DocReviews by Document Type

Category	*n*	*total words in documents*	*total words in commentary*	*effectivity*
Type 2	10	1302	696	0.535
Type 3	58	27636	3181	0.115
Type 3 w/o minutes	8	4433	909	0.205
Type 4	10	8581	2914	0.340
All Types	78	37519	6791	0.181

The DocReviews that received comments were analyzed and two columns were produced by database query: size of the base document and the summation of the size of the commentary on the DocReview. This table was imported into the spreadsheet. For each DocReview, the size of the commentary was divided by the size of the base document to yield effectivity. A column was created for the effectivity. An XY scattergram was produced with segment size on the X-axis and effectivity on the Y-axis. Five effectivity distributions were studied: all DocReviews by document type, meeting minutes (most of the type 3 documents), and all DocReviews less the meeting minutes.

Studying the distributions of the three types shows three very distinct populations, type 2 with very strong logarithmic decay of effectivity with increasing base document size, type 3 documents with a very low effectivity and an almost random distribution (see Figure VI), and type 4 with logarithmic decay of effectivity. Considering the strong (R2

= 0.4416) logarithmic decay of effectivity with increasing base document size seen in Proposition B1 (§5.1.6.3), the nature of type 3 documents needs to be examined more closely.

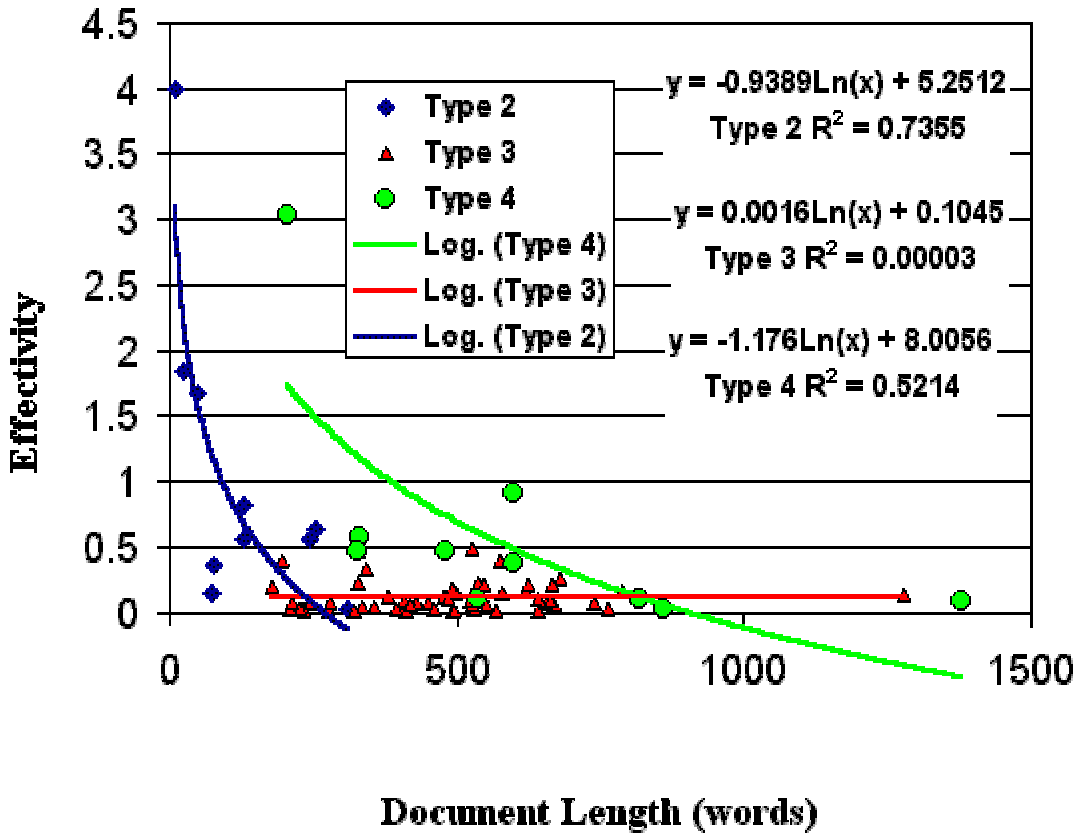


Figure VI Effectivity to Document Length by Type

Table XI Regression Analysis Summary

Type	df	F	P_{slope}	$P_{intercept}$	R	Std Err
2	9	22.2	0.0015	$3.3x10^{-8}$	0.858	0.593
3	57	0.001	0.966	0.644	0.0057	0.117
4	9	8.72	0.018	0.013	0.722	0.658

Type 3 documents are working drafts, in the data examined here either position papers submitted for a workshop or minutes of weekly group meetings. Meeting minutes are a highly stable and consistent genre that does not attract much discussion, unless discussion topics were not reported or were reported incorrectly. All the meeting minutes were consistently formatted and prepared by only three people. They were separated from the position papers and examined and the effectivity was found to be essentially randomly distributed (R = 0.05) with respect to document length (see Figure VII). With 49 degrees of freedom a value of F = 1.21 was found. The slope was positive with P = 0.730. The P value of the intercept was 0.033.

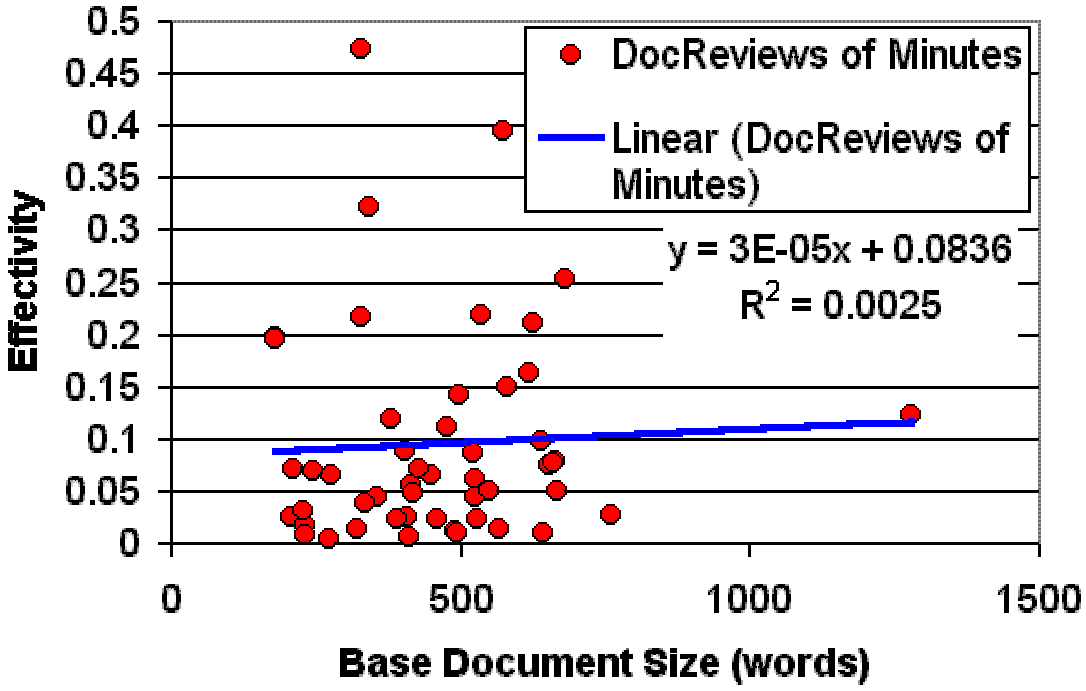


Figure VII Effectivity of DocReviews of Minutes

Based on the finding that meeting minutes formed an essentially random cluster of data points that was well distributed at the knee (document size 200-800) of the logarithmic regression line, it was decided to plot all DocReviews except the meeting minutes. This distribution contains documents (n = 28) that are more likely to stimulate substantive dialog.

A correlation of 0.714 on the logarithmic regression line confirms a strong negative logarithmic relationship. With 27 degrees of freedom a value of F = 27.1 was found. As expected the slope was negative with $P = 1.98 \times 10^{-5}$. The P value of the intercept was 1.7×10^{-6}.

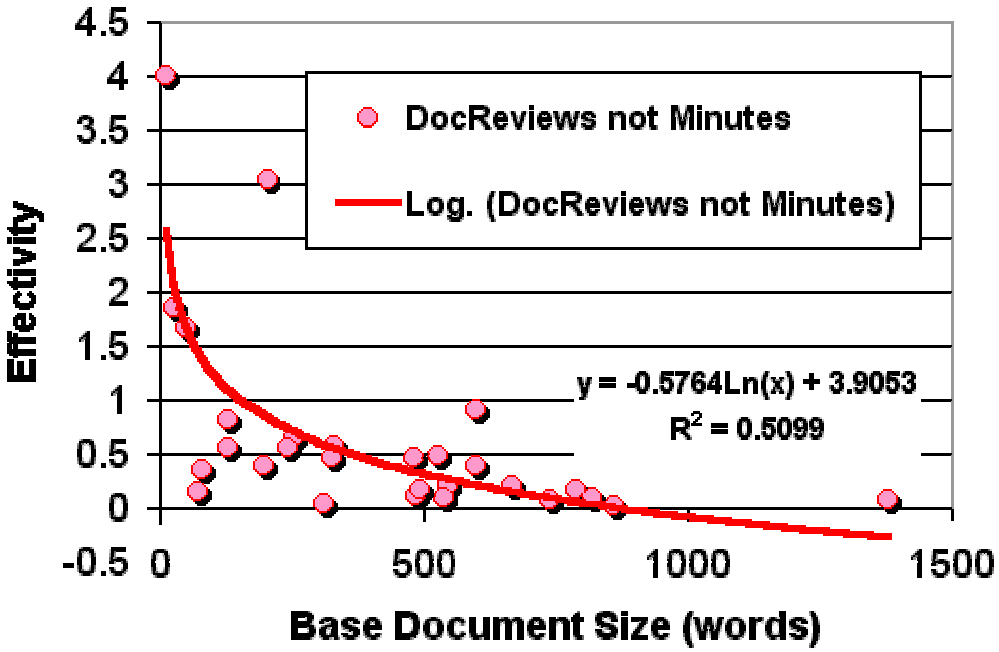


Figure VIII Effectivity of DocReviews not Minutes

Discussion of Findings:

The hypothesis is soundly rejected. It is clear that less finished documents attract more participation than do more polished documents. This is likely due to the presence of more opportunities for change through collaborative critique.

5.1.6.8 Proposition D2: The nature of social commentary will vary with the type of document.

It is expected that the more formal nature of higher quality documents will evoke a more formal commentary as opposed to the informal and preliminary nature of the less mature documents.

Operationalization:

The social character of the comments is operationalized as the distribution of the Bales codes categories for each of the document types. The Bales Interaction Process Analysis categorizes all speech acts, including gestures, into twelve codes. Many of the Bales codes are specific to face-to-face dialog, so we must eliminate those codes in order to make a comparison. Bales grouped the twelve codes into four categories that are generic and form a good basis of comparison. These categories are: Social-emotive area: positive (positive reactions), Task area: positive (problem-solving attempts), Task area: negative (questions), and Social-emotive area: negative (negative reactions). The central two categories are further generalized into a supercategory of the task area, while the extremes are generalized into the social-emotive area.

For each of the four Bales categories, the percentages of commentary codes by document type (n = 3) are graphed.

Data conditioning:

None.

Data Analysis:

The Bales category distributions of DocReview annotations by document type demonstrate that the annotations are almost never negative reactions. The annotations that show positive reactions are more often directed to the more finished documents (type 4) than to the working and rough drafts (types 3 and 2). Questions are asked over twice as often in type 2 (rough) documents as in type 4 (finished documents).

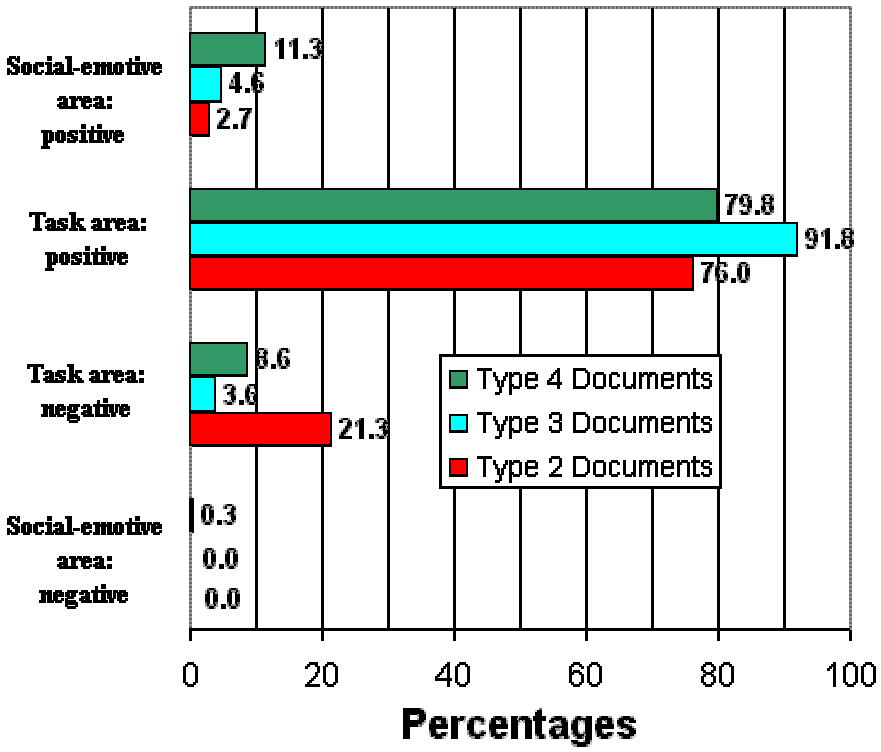


Figure IX Distribution of Bales Categories by Type

We find that the null hypothesis that there will be no difference in the Bales category distribution between document types can be rejected. With six degrees of freedom, Chi squared = 46.5. This result is significant at <0.000001.

Discussion of Findings:

Finished documents are viewed more positively than rough documents in DocReview. Most commentary is directed toward problem solving.

5.1.6.9 Proposition D3: The nature of substantive commentary will vary with the type of document.

High quality documents such as Research Web Essays (type 4) will attract relatively few negative comments, just because the documents are likely to contain few errors and omissions. On the other hand speculative documents (type 2) are likely to attract negative commentary due to their incomplete and unfinished nature. Working documents are likely to occupy an intermediate position.

Operationalization:

The substantive character of the comments is operationalized as the distribution of the Meyers structurational argumentation codes categories for each of the document types.

Data conditioning:

None.

Data Analysis:

Of interest is the distribution of reinforcer percentages among the types of DocReviews. The more polished (Types 3 and 4) documents draw over twice the percentage of reinforcers than do the rough (Type 2) documents. This is distribution is inversely mirrored, weakly, by a corresponding presence of a lower percentage of promptors in the polished documents as compared to the rough documents.

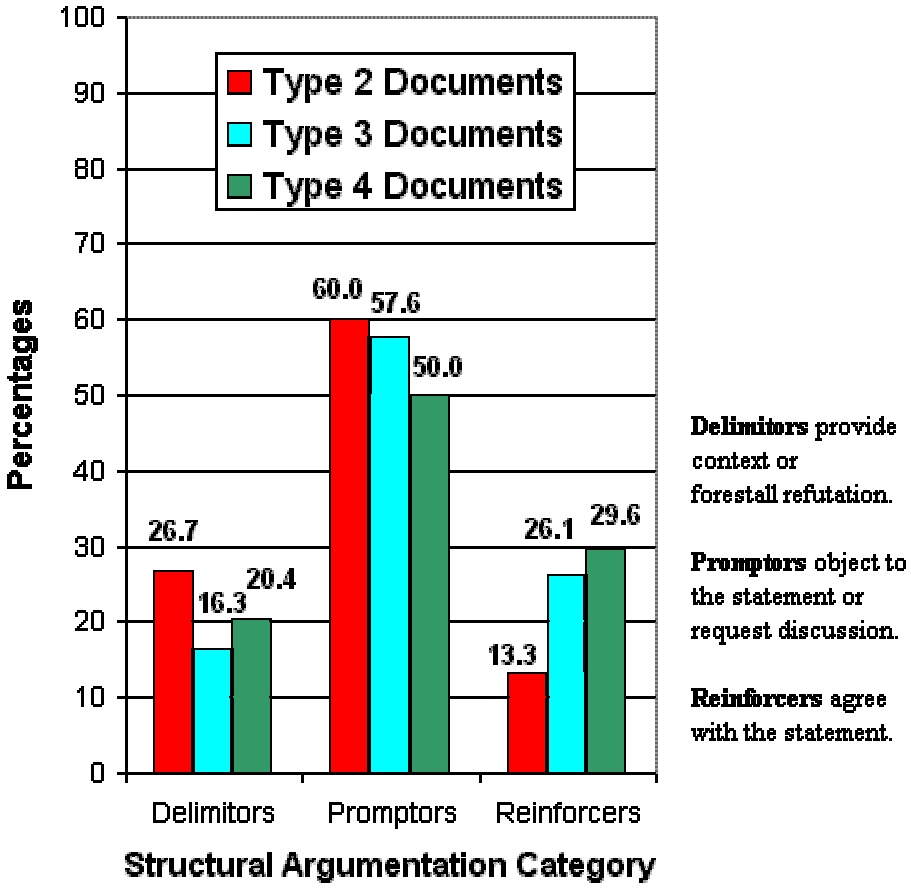


Figure X Substantive Commentary by Document Type

We find that the null hypothesis that there will be no difference in the Meyers Argumentation Code category distribution among the document types can be rejected, but only very weakly. With four degrees of freedom, Chi squared = 3.92. This result is significant only at <0.5.

Discussion of Findings:

The distribution of argumentation categories is only weakly contingent on document type. There are indications that polished documents will attract more agreement and somewhat fewer objections than rough documents.

5.1.6.10 Other Findings

Exponential decay of multiple comments is seen. The regression line shows a correlation of 0.941 for classes of comment counts, 0 to 6.

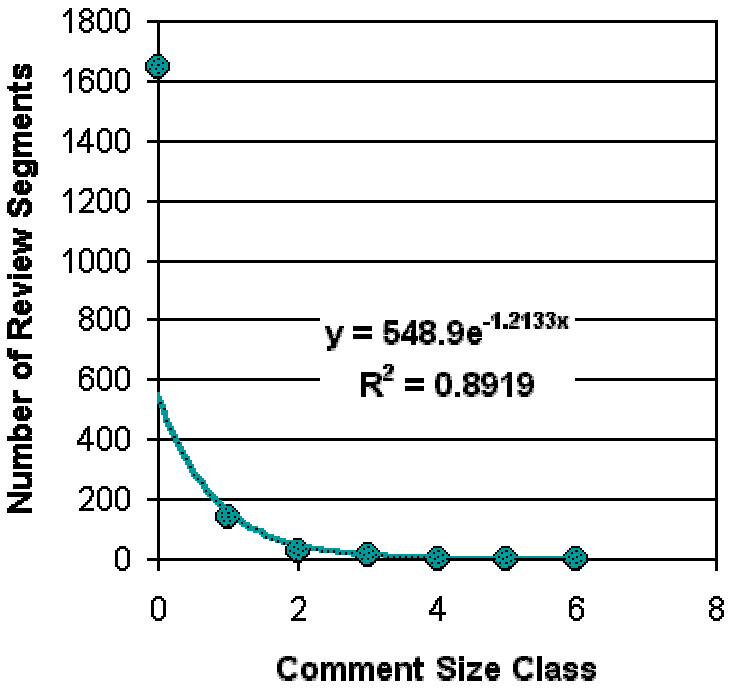


Figure XI Exponential Decay of Comment Counts

5.1.7 Conclusions

The social character of the dialog elicited by DocReview shows substantial departures from face-to-face dialog [prop A1]. The social character of the dialog is much less emotive than face-to-face dialog and the task oriented dialog is very clearly oriented toward problem solving, with questioning being only slightly less than in face-to-face.

The substantive nature of dialog in DocReviews [prop A2] is very concentrated in constructive disagreements with the statements in the DocReview. Conversely, agreements are much less frequent than in face-to-face dialog. Most of these agreements include amplifications. This finding reinforces the similar findings in the study of the social nature of the dialog [prop A1].

Findings related to the size of the base document and the segment size found that the effectivity of the DocReview decays logarithmically with increasing base document size [prop B1]. Commentary size is directly, but not strongly, proportional to segment size [prop B2]. The effectivity of a review segment shows logarithmic decline with increasing segment length [prop B3]. This finding indicates that the document segmentation strategy should avoid long segments.

Analysis of the descriptive statistics on the document size shows that the length of annotations is significantly longer in more finished documents (type 4), perhaps reflecting the willingness to spend more time on "serious" documents, and shortest in working documents (type 3). Annotations on rough documents (type 2) fall into an intermediate length class, perhaps because they need more work to bring them to acceptable quality.

Comparing DocReview to roughly comparable products shows that no important features were overlooked in DocReview, though no product has implemented the features just as DocReview has [prop C1]. This convergence of design demonstrates that DocReview's design is in the mainstream. The differences in design implementation are largely due to differences in audience and commercial aspirations.

An attempt to measure the effect of base document quality on the effectivity (the ratio of words of commentary to words in the base document) of the DocReview found [prop D1] that (with exceptions) effectivity of documents declined with increasing quality, corroborating the findings of prop B1. Measuring the effect of base document quality on the social nature of the dialog showed comparable distributions among the Bales categories [prop D2] in all document types. The minor differences speak perhaps more to the consistent categorization of documents than to the significance of the differences. In the case of substantive dialog (Meyers codes), similar comparable distributions were seen [prop D3]; however there was an apparent, but insignificant, increase in agreements (reinforcers) with increasing quality. A corresponding decrease in objections was also seen.

5.2 Case Studies of Research Webs

There have been four attempts to develop Research Webs, and while none were successful, much knowledge was gathered. The design of Research Webs was largely theory driven, but practical experience has been gained from the attempts to develop RWs. Much of the knowledge was applied to the tools, especially DocReview, but other knowledge gained has come from understanding human nature, especially the failure to participate.

While perhaps a few minor problems with RWs may be laid at the feet of technology, by far and away most barriers are related to human behavior: psychological, sociological, organizational, and cultural. The research questions we examine are designed to uncover behavior patterns and the reasons that may contribute to causing those patterns. Discussion of counterproductive behavior leads us to some suggestions to modify that behavior. To the extent that such causes may be remediated by technology, the tools applied to the RW will be augmented or changed.

5.2.1 History of Research Web Technology

Technological aspects of Research Webs have co-evolved with the World Wide Web. Beginning from the initial release of the Mosaic web browser in 1993, the potential of the WWW was clear. The technology was at that time blooming on a monthly basis and, as time allowed, the new capabilities were incorporated into the Research Webs.

The prototype RW was the Migration RW, originally a simple hypertext installation with but one image, a Dorothea Lange photograph of an Okie family on the road to California during the Great Depression. The first new feature incorporated was the inclusion of an organizing model, a diagram of the migration process from information gathering to settlement. The diagram, a petri net[34], was image mapped so the user could click on an object in the diagram, a node or link, and be transferred to a page that described the object. This RW was essentially a technical proof of concept site, and included crude bibliographic and glossary links, and a commenting capability limited to global comments on each web page. No research team was assembled.

The first RW, the Chromium VI RW, was the first opportunity to engage a team of scholars. By that time DocReview had been developed, allowing participants to make annotations on small chunks of text from the web page. DocReview was used to annotate documents and meeting minutes. In the later stages of that RW, JavaScript was employed to provide the ability to pop up small auxiliary windows on the screen. If the user wanted to view a DocReview, bibliographic information, a glossary definition, footnote, or a sidebar, those features would be displayed in small windows without overwriting the main document window.

In the final two RWs, the Soil Crust RW and the Earthquake Disruption RW, the ability to annotate bibliographic entries and glossary entries was added, bringing the RW technological environment to its current state.

5.2.2 Research Questions

There have been several research efforts that used the collaborative tools of the Research Web environment. Four of these efforts utilized the fully developed concept of the research web, while others used only parts of the concept or the incompletely developed concept. The case study of the Research Webs will address several research questions.

The research questions and associated propositions are:

1. What was the focus (issue domain) of the RW?

- A diffuse focus for the RW will destroy it.
- Confederations (groups with different focuses) under a single RW will fail.

2. What were the geographic distribution effects on the RW?

- RWs with strong concentrations of people who can easily communicate in person will fail.

- RWs with widely distributed members who live in different time zones are more likely to succeed than RWs with concentrated membership.

3. How many people were invited to participate in the Research Web?

- The critical mass theory holds for research webs.

4. What incentive(s) did each of the participants have to participate?

- In order to be successful the RW must provide rewards beyond authorship.

While several additional questions were considered, only these could be addressed properly. The questions were posed after the active lives of the Research Webs. Several could not be answered due to lack of data. Others would have needed questionnaires for proper evaluation. Several of these currently unaddressable questions are presented in the Future Research section.

5.2.3 Design of Data Collection System

The data collected on the research webs consists primarily of copies of the web sites, meeting minutes, and correspondence between the RW's scientific coordinator and the facilitator. Web sites include data not only in web pages, but also data in the form of annotations in DocReviews, Annotated HyperBibliographies, and Annotated HyperGlossaries. DocReview builds a log file, which contains all transaction activity: creation, annotation, reading and archiving. There were two hosting servers that went off-line during this research, but the four Research Webs described below were captured

before destruction, or were recovered from the server host after the server software failed.

5.2.4 The Research Webs

In this section we will describe four cases, a prototype and three Research Webs, and using those cases and events and circumstances in associated enterprises will discuss the research questions (above §5.2.2). Each RW had as its issue domain a topic that was subjected to scholarly research. All the cases were hosted on web servers at the University of Washington and were facilitated by the author.

5.2.4.1 Migration Prototype Research Web

This Research Web was initiated as a prototype and test bed for the Research Web concept to demonstrate the power of the WWW to facilitate research. The topic was internal migration viewed from a behavioral standpoint. The site capitalized on the work done in my Master's Thesis, a section called "A Model of the Migration Process"[35]. The importance of this site is related to the testing of the technology later applied to Research Webs and to the realization that the Research webs were a social organization driven by social and personal goals, not technology.

5.2.4.2 CREAT and The Chromium VI Research Web

The Collaborative Risk Evaluation Approaches Team (CREAT) was an attempt to build an interdisciplinary team to investigate a small set of problems centered on the cleanup at the Hanford Nuclear Reservation. It was staffed with members who were supported by the Consortium for Risk Evaluation with Stakeholder Participation (CRESP) project at the University of Washington. CREAT was the first user of many of the tools now used in Research Webs. A Research Web was initiated to investigate one issue that concerned CREAT, hexavalent chromium contamination. It was hoped that the Chromium VI RW would serve as a template for describing several other contamination issues.

5.2.4.2.1 Mission and RW Topic

> *The mission of the Collaborative Risk Evaluation Approaches Team (CREAT) is to provide information about hazards and risks to human health, ecological health, economic health, and socio-cultural health within and around DOE sites. CREAT is developing an easily accessible, computer-based tool for collecting, cross-indexing, displaying, and comparing the components influencing risk. The tool will enable interested parties to understand what is at risk and how and why it is at risk. In addition, it will provide an analytical means for comparing risks within one site as well as across sites within the DOE complex. Finally, it will bring the results of research by members of CRESP to the attention of stakeholders, Tribes, DOE and other interested parties and will provide a forum for discussion of important issues in the area of risk evaluation and hazard reduction.*
>
> --- CREAT mission statement produced by the team, 1996

It was decided to open a Research Web on the topic of environmental impact and the remediation of hexavalent chromium contamination. The mission statement of CREAT's RW was, "Initially, the focus of this research web is to examine one specific hazard: hexavalent chromium found in the 100 Area [near the plutonium production reactors] along the Columbia River on the Hanford Site. The principal risk posed by this hazard is as a stressor to the salmon stocks that spawn in the Hanford Reach of the Columbia River. There are some risks to human health, and these aspects will be investigated as well. The research web will work with absolute risks not relative risks. By using this approach we hope to develop a conceptual and informational framework that can be applied to other hazards."

5.2.4.2.2 Organization

CREAT was always an ad hoc voluntary organization. It was never funded on any budget, but was a management-approved activity. Its existence was justified by two possibilities: that it might produce publishable research; and that it would produce

explanations, "fact sheets," of detailed issues surrounding some of the health hazard problems from the contamination at the Hanford site.

The nominal leader of CREAT was a member of the CRESP administration and the administrative leader was a scholar funded as a member of a CRESP task group. There were six contributors of Research Web Essays, and twelve team members contributed annotations. An additional four team members participated in face-to-face meetings, but not in the RW. The team operated as a collaboration, with all documents open to annotation from all members.

5.2.4.2.3 Focus

CREAT emerged as a focus group within the Health Hazards Identification Group of CRESP. The group was founded in September of 1996 as the Health Hazard Identification Focus Group (HIF). Briefly this group was identified by its nature and methods as the "inter-disciplinary collaborative risk evaluation and analysis group" (ICREA). The group took the name CREAT on October 22, 1996.

The founding challenge for the CREAT group was to produce short essays on ten questions on hexavalent chromium contamination. Three members of the ICREA GROUP proposed the questions. At my urging, the CREAT group enthusiastically agreed to participate in a Research Web. The ten questions were placed on the RW's web site on December 6, 1996 and were refined by members of CREAT and other members of CRESP. The ten questions were then answered by short essays by team members, and were put into the RW web site for viewing and for critical annotation using DocReview.

The ten questions were:

What is Chromium?

How did Chromium become a contaminant?

Why is Chromium a "contaminant of concern"?

Where is the contamination and how much is there?

How do the levels of Chromium compare with regulatory standards?

How is the Chromium concentration measured?

What is the quality of the Chromium data?

How have Chromium concentrations changed over time?

What is presently being done to mitigate or control the Chromium hazard?

Have new methods of controlling the Chromium hazard been suggested?

In addition to essays on the ten questions, seven other essays were contributed to the site. One team member created four ecologically oriented essays dealing with bioremediation of hexavalent chromium and the effects of hexavalent chromium on salmon, invertebrates, and aquatic plants. I introduced three additional documents in order to introduce some measure of process modeling into the site. These documents were essays on the Environmental Remediation Disposal Facility (ERDF), the pump-and-treat decontamination processes practiced at Hanford, and a process model of the chromium contamination processes (See Figure XII below). The process model was designed as an organizing model to tie together several of the other essays.

CREAT's work was extended to a parallel study of tritium contamination, but this work never reached the Research Web stage. The efforts of the CREAT team members were directed by management toward a grander project called the Risk Information Tool (RIT) that slowly became moribund due to a number of issues: lack of resources, competition between CRESP management units for control of public information, lack of participation from most CRESP task groups, and a diversion of effort due to a management mandate for concentration on scholarly publications. Unfortunately for the Research Web, RIT was designed for a different audience than the chromium RW, so there was an incompatible conflict in goals. While RWs could provide information to the RIT, it was RIT that had a wider audience, and because of that, RIT received the attention of CREAT.

The Chromium VI RW was generally considered to be a good idea, but was destroyed by incorporation into RIT, an enterprise that languished. Its life, from inception to inactivity, was about ten months.

5.2.4.2.4 Case Study Data

Meeting minutes were recorded and archived, most in DocReview. In addition to CREAT minutes, minutes of the Data Characterization, Analysis and Statistics (DCAS) task group of CRESP recorded some CREAT activity, since several DCAS members were CREAT members. Personal e-mail archives also contain discussions of issues relating to CREAT.

5.2.4.2.5 The Web Site

There were seventeen essays on the site, the ten questions and seven others. Members contributed four essays on ecological topics, and the facilitator contributed three more to

provide an organizing model (see Figure XII below) for the RW.

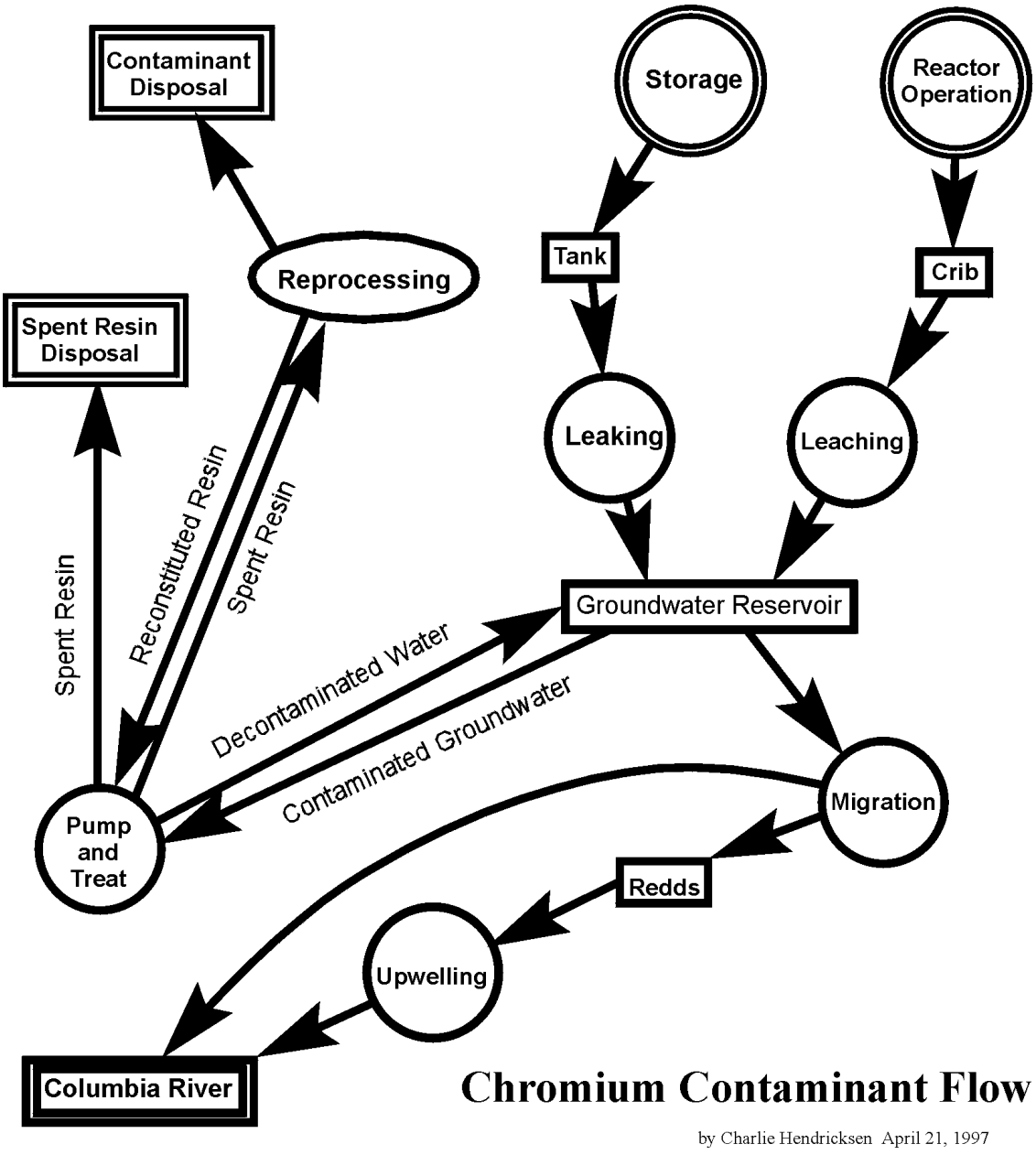


Figure XII Organizing Model for the Chromium RW

5.2.4.2.6 Conscription Devices

The seventeen essays were all DocReviewed and most received annotations. The process model was not image mapped, and because of that only general comments about the process could be made. There was no Annotated HyperBibliography or Annotated HyperGlossary, as that software was not yet in existence.

5.2.4.2.7 Participation Profile

Most members of CREAT, especially the content providers, were well engaged and participated in DocReviews. Several members of the team were frank advocates. Requests and reminders to review documents were issued freely, but consciously avoided importunity. Four members of the group never participated online, but did attend meetings. One of these was openly resistant to the Internet, and found hours of time to attend meetings but never the minutes to participate online. Participation by the team would likely have continued, but a motivating stream of new content (conscription devices) never developed.

5.2.4.3 Soil Crusts RW

The Microbiotic Soil Crusts RW was begun in late 1996 with the intention of creating a Research Web specializing in the study of crusts of lichen, mosses and cyanobacteria that form on some soil surfaces of semi-arid lands. The topic was a natural extension of the scientific coordinator's research interests, and would also provide a collaborative environment for the study and cataloging of soil crusts on the Hanford Nuclear Reservation in Washington State.

5.2.4.3.1 Mission and Topic

This Research Web was founded to coordinate the efforts of eight researchers that were

studying microbiotic soil crusts in the northern steppe ecotone. The work that brought them together was a survey of soil crust lichens on the Hanford Nuclear Reservation. There was a need to obtain this data in order to determine the feasibility of using these micro-communities for indicators to measure disturbance of the ecological habitat on the Reservation.

5.2.4.3.2 Organization

This enterprise was funded indirectly by CRESP and Pacific Northwest National Laboratory (PNNL), and by a small grant from The Nature Conservancy. CRESP and PNNL have ongoing research interests at Hanford. The Nature Conservancy was a contributor and interested party by virtue of its long research association with the Fitzner-Eberhardt Arid Lands Ecology (ALE) Reserve. Team members were from the University of Washington (2), Washington State University, The Nature Conservancy, and Pacific Northwest National Laboratory (2).

5.2.4.3.3 Focus

The RW was begun in late 1996 on the heels of the successes of the CREAT Research Web. The scientific coordinator was one of the energetic collaborators in the CREAT Chromium RW. There were high hopes for this RW as it included a set of collaborators with an authoring project, funding, an enthusiastic scientific coordinator interested in long-term research, and an eager facilitator.

A paper was to be produced by the team as the result of field work performed by seven members of the team. The principal author was somewhat reluctant to share early drafts of the manuscript with the team at large, even by hardcopy; but eventually a draft of the

paper was DocReviewed on the RW. Another paper was planned, on field sampling techniques, by the scientific coordinator, but never reached first draft stage.

The Research Web slowly became moribund in 1998 as its only effective conscription object, the scholarly paper[36], neared completion. The scientific coordinator and facilitator attempted to keep the RW alive in hopes of attracting interest among the very small and specialized soil crust community. Closing the server computer in 2001 terminated all activity.

5.2.4.3.4 Case Study Data

The complete web site and 344 e-mail messages related to the RW have been archived.

5.2.4.3.5 The Web Site

The web site contained a full complement of tools: a home page, essays, an interactive page to discuss research interests, an interactive page designed to discuss plans for Autumn 1997 field work, Annotated HyperBibliography with author and title indices, species list, a Lexicon installation, Annotated HyperGlossary, a photo album, indices to DocReviews (both active and archived), both public and team partitions, and an authoring team. Minutes of team meetings were mounted on the web site and were DocReviewed.

The facilitator, with the help of the scientific coordinator, planned and prototyped a new tool, the Species List. This software was designed not only for cataloging each of the lichen species found, but also to provide an online annotation capability so specialists could insert "micro-essays" on a species. These "micro-essays" called Specialist Views

in the species list were designed to allow the appending of interdisciplinary knowledge to the species list. For instance, specialists in range management or fire ecology might have notes on species important to their work. This tool was an example of creation of a new tool suited to the distinctive needs of the team's research.

5.2.4.3.6 Conscription Devices

Conscription devices installed to draw out the knowledge of the team members included a Lexicon with many entries intended to assemble entries for the Annotated HyperGlossary, an Annotated HyperBibliography with 78 entries, two essays, a species list prototype, 7 DocReviews including a draft of a professional paper, and an interactive research interest page. There was no organizing model presented.

5.2.4.3.7 Participation Profile

The scientific coordinator promoted the RW actively. The facilitator worked closely with the facilitator, but engaged with the team only at the request of members, making no independent contacts. The collaborators were well distributed among four organizations in Washington State. The distance from Seattle-Tacoma to Richland was such that face-to-face meetings would require considerable travel time by road or air. At best, a meeting would take one very long day of travel. It was assumed that this distribution would encourage on-line participation.

Two invited members never participated in the RW, and one of those, a senior researcher who would have been of great value to the team, could not be convinced of the utility of the WWW, and absolutely refused to interact with materials not in hardcopy. The attitude of this person to the technology was so extreme that they were stereotypical. Attempts to recruit new members were made in 1997, but little interest beyond polite

replies was encountered. Participation by members other than the scientific coordinator ceased after the first draft of the paper was DocReviewed.

The RW was presented at the 1998 convention of the Society for Ecological Restoration ("Can soil crusts act as indicators of the biological condition of the shrub-steppe? Using the world wide web to foster scientific collaboration."). It was received with interest and a few people expressed a desire to participate in such an enterprise, but nothing developed.

5.2.4.4 Earthquake Disaster Mitigation

In the Autumn of 1998 work was begun on a RW to support the US-Japan Cooperative Research on Urban Earthquake Disaster Mitigation. This project was based on the experiences of researchers with two major earthquakes in the early 1990's: the Northridge Earthquake in Southern California and the Great Hanshin Earthquake in Kobe Japan. Damage to regional transportation infrastuctures was profound in both these disasters.

5.2.4.4.1 Mission and Topic

> *This WWW site supports an interdisciplinary team of scholars studying the impact of catastrophic earthquakes on urban transportation systems. This team is distributed around the Pacific Rim in Japan and the United States. The team's goals are to develop both a broad synthesis of the impact of large earthquakes on transportation systems and several more specialized studies. The knowledge produced by the specialized studies is expected to support the broad synthesis; and the synthesis is expected to illuminate the specialized studies and to produce new insights and hypotheses.*
>
> --- from "Site Design and Research Support" web page

5.2.4.4.2 Organization

The team was composed of twelve scholars well dispersed on the Pacific Rim: one in Seattle, five in Los Angeles, and six in Japan. The Japanese scholars were all affiliated with the Disaster Prevention Research Institute (DPRI) of Kyoto University. The American scholars were not organized into a formal association. The scientific coordinator was a colleague of the author at the University of Washington.

5.2.4.4.3 Process

My colleagues were always on the lookout for research enterprises that might become Research Webs. Tim Nyerges, my committee chair, told Stephanie Chang about the concept and how it had worked in practice. She contacted me in September of 1998, and the decision to go ahead with a Research Web was made in the following month. At a team meeting in December, Dr. Chang presented the Research Web concept and it was accepted with enthusiasm.

Content in the form of meeting minutes, CVs, and professional papers were added over the year of 1999, but participation was minimal. In March of 2000, a public partition was

added, but that did not spur any further participation. At the time of writing, the RW is moribund.

5.2.4.4.4 Case Study Data

The complete web site is available and 163 e-mail messages related to the RW have been archived.

5.2.4.4.5 The Web Site

There were 40 web pages in the RW web site. The site was partitioned into two partitions, a public partition with 14 pages and a passworded team partition with 26 pages plus DocReviews.

In the team partition, the web site contained several infrastructural pages: a team page, listing all members and their affiliations, with links to the five members who have home pages; a Mail Room page that allowed the user to send e-mail to any team member; a What's New page that allowed the user to obtain a list of activity in any or all DocReviews; a meeting schedule page; a page describing the web site and the support available for the researchers; a page with links to DocReviewed professional papers; an index to all DocReviews; a page providing links to transportation system data in Seattle, the region of interest to the scientific coordinator; and links to project archives.

5.2.4.4.6 Conscription Devices

Two published papers were mounted on the site in DocReview format so the team members could annotate the documents. No annotations were made. Eleven documents

were DocReviewed and collectively gathered only seven annotations, all by the scientific coordinator and one colleague. There was no organizing model, Annotated HyperBibliography or Annotated HyperGlossary developed for this RW.

5.2.4.4.7 Participation Profile

Of the twelve researchers, only two participated in the Research Web. The scientific coordinator promoted the RW at appropriate times, avoiding overt advocacy. The facilitator engaged only at the request of members, making no independent contacts.

There was a lack of conscription devices to attract participation. Working papers were not shared except within the authoring teams. Two papers were presented for DocReview, not as working papers, but as finished or submitted documents. One computer model was produced, Walter Svekla's master's thesis, but was not incorporated into the Research Web.

5.2.5 Case Study Analysis

In this section, the data gathered from experiences with the four Research Webs are analyzed. Several research questions are introduced and interpreted, and the propositions arising from the research questions are discussed.

5.2.5.1 Research Question 1:

What was the focus (issue domain) of the RW?

Framework for Analysis:

The issue domain of the RW can be sharply focused or very diffuse. Even a sharply focused issue domain may be, however, far too large in scope for a RW team. An

expansive issue domain may be so sparsely populated by content that the units are not interdependent. Independent units will not encourage mutual collaboration by their authoring teams. In other words, they will not be contributing to a unified whole. There is also the possibility that the scope of the RW is too narrow for a long-term collaborative enterprise. Such is the case where the collaboration is focused on a single scholarly paper with no interest in either enlarging the scope of the RW to include closely related topics, or elaborating the details of the objects or processes of the RW; in other words, by building neither a supermodel nor submodels.

Propositions:

I) A diffuse focus for the RW will likely result in little participation.

An organizing model that provides a central point of interest and a set of potential work objects. The function of work objects is not only to produce tangible results and publishable documents, but also to conscript the members of the team into active participation. The principal conscription devices are models of the issue domain. Other conscription devices, such as RW essays, bibliographic information, glossaries, discussion forums, and document reviews, encourage participation; but such participation is generally not central to the research but rather to discrete initiatives within the research effort.

Focus and scope are closely related. Focus is the issue domain at the core of the RW. Scope is the outer boundary of reasonable extensibility of the RW. A narrow focus is necessary to provide interdependency of the topics of the issue domain. If the authoring teams address unrelated topics, then it is likely that the research team will self-segregate

into subteams. The volume of communications between the subteams will be minimal. Indeed, the division of effort may cause the subteams to drop below critical mass. An overly narrow scope will not provide sufficient research opportunities for a research team that is large enough to provide critical mass.

Chromium VI Findings

The issue domain of the RW was the universe of causes, processes, damage, hazards, and remediation efforts associated with Chromium VI chemical contamination from the plutonium production reactors along the Columbia River in the Hanford Nuclear Reservation. A process flow model provided the organizing model. This model related many of the physical objects and processes defining the behavior and character of the contamination. Each element of the organizing model provided ample scope for expansion of the topic by description and theory building. Many of these elements provided topics for potential research papers including the application of geology, natural history, engineering, ecology, and human health to the issue domain. Indeed, plans for research papers were begun on bioremediation and the effects of contamination on salmon spawning; and descriptive pages, such as one that described the Environmental Restoration Disposal Site, a repository for contaminated soil, were added to support the model. The life of this RW was not sufficiently long to determine if the organizing model would have been effective.

This was the first RW and as such all its aspects were not well understood by the participants. The organizing model was not criticized and little attention was given to contribution of essays to flesh it out. What *was* well understood was that ten questions needed to be answered in short essays. Those ten essays occupied the attention of many of the members. Two members did contribute essays beyond the ten questions. This RW had proper scope, focus and a coherent organizing model. Its ultimate failure was not

due to shortcomings in scale or focus, but rather to changes in the mission of the research team.

Migration Findings

The issue domain of the prototype RW is migration by household units focused on the behavioral processes by which such migration comes about. This issue domain is far too broad for a RW as proposed in this dissertation. Its organizing model is quite adequate for the description of process and ultimately for theory building. Very likely, even the next lower level of abstraction is such that it will also be of interest only to scholars with broad synthetic theory building interests. Only at even lower levels still less abstract will issue domains be found that are properly focused for Research Webs. In support of the contention that such very high levels of abstraction preferentially attract theory building is the observation that three well-established scholars contributed (Tobler, Davis, Amrhein). Despite a personal appeal to a graduate seminar on migration, little interest was shown and no substantive contributions were made.

Soil Crusts Findings

The issue domain of the Soil Crust RW was the nature of the soil crusts in the northern shrub steppe biome. Soil crusts are mats of lichens, mosses, and cyanobacteria that form on undisturbed soil surfaces in semi-arid lands. The scientific coordinator and the facilitator agreed that the topic was sufficiently specialized to provide a proper scope for a RW. Given those assumptions, work was started immediately on a comprehensive bibliography and glossary of terms. There was no organizing model, though there were several bases for organizing models. The ecology of soil crusts is characterized by mutualism, symbiosis and perhaps parasitism. How an ecosystem dominated by only four classes of organisms: cyanobacteria, lichens (fungus and algae), and mosses

interacts as a general system could serve as an organizing model. Microclimates and geographical models of the ranges of species are other potential organizing models.

The project's immediate goal was to provide a forum for the cataloguing of the soil crusts on the Hanford reservation, with the intent to expand the scope of studies in the future. Once the principal work object, a research paper cataloging the lichens, was finished participation dried up. Several conscription devices were available, but did not attract participation from anyone except the scientific coordinator. The fact that this properly scoped RW went moribund is a demonstration of the importance of rewards. When the reward of authorship had been spent, there was no further participation. If there had been an organizing model, a plan for a series of research articles, and scholars to perform the research, the RW might have survived.

Earthquake Disruption Findings

The issue domain of the Earthquake Disruption RW was the impact of large earthquakes on urban transportation systems. The grant proposal that described the work clearly indicated that the work to be done was a reconnaissance of the field designed to support scholars who would hopefully describe the field and generate hypotheses. The diffuse goals of the work precluded any organizing model, though the mandated synthesis of the field will perhaps provide one.

With no organizing model and a working mandate that encouraged independent, though related research, there were few common work objects. Furthermore, all rewards were clearly related to production of research papers. Potentially unifying conscription objects such as glossaries and bibliographies were not initiated. This RW had no research focus, thus little need for research collaboration outside the authoring teams.

Discussion of Findings:

A research web needs to be properly scoped to be successful. The scope of the RW is determined in part by topic and in part by size. The number of members required to perform the research determines the size or scale of the RW. The permanent members of the research team must be sufficiently interested in the entire issue domain to contribute criticism to all documents. Clearly there needs to be a critical mass of dedicated researchers; beyond that there is support staff needed, including at least a scientific coordinator and a facilitator. The topic, or issue domain, must cover a set of clearly interrelated elements that are sufficiently specialized. The specialization should be such that each element is either a good topic for a single research paper or a topic that can encompass a small family of very closely related research papers.

Scope was clearly an issue in the lack of success with the Migration RW. It was so vast that only philosophers could work at that level. While the organizing model was very interesting, the elements of that model were not sufficiently specialized to produce either research papers or small families of research papers.

A lack of interrelation is indicated by difficulty in developing an organizing model. The Earthquake Disruption RW showed this difficulty. Likely the work undertaken in the founding grant will act as a research reconnaissance and will synthesize organizing models for the issue domain. The Soil Crust RW also had no organizing model because the RW was organized around a single research paper. Had the team been interested in developing the RW around biological or ecological system models rather than a cataloging of species present in a given area, it may have survived.

Both proper scope and focus were demonstrated in the Chromium VI RW. Lack of success can be attributed to a failure of leadership. While management initially approved the RW, it failed to recruit specialists to contribute research essays (and eventually papers). The RW ended when management redirected the efforts of the research team into a project focused on providing information rather than supporting exploratory research.

II) Confederations (groups with different focuses) under a single RW will fail.

The CRESP project that supported two of these cases was formed as a large-scale collaborative project joining several disciplinary specialty groups with a single support group dedicated to gathering data, providing statistical services and supporting a Geographical Information System. While CRESP was far too large for a single RW, it could easily have provided an umbrella organization to support several loosely interrelated RWs. Instead it devolved into a successful confederation of independent authoring groups that generated many professional papers, but little collaborative work toward its original mandate. Failure to provide an organizing model resulted in pathologies that ended all collaboration. Not only did each disciplinary group stay within their specialty, but also the existence of two cooperating Universities locked into an inferior/superior structure took its toll. The research team was fragmented both by discipline and institution. These same pathologies can destroy RWs.

Chromium VI Findings

The Chromium VI RW had only a single group of collaborators (CREAT), but their efforts were fragmented in several directions. First, the members of CREAT were participating under a matrix management agreement, essentially on part-time loan from several task groups of the CRESP project. While this regime was interdisciplinary, and thus positive from a collaborative sense, there was competition in agenda setting. The research direction could remain interdisciplinary, or could veer off into a specialty area

such as ecology, toxicology or human health hazards. Specialists could satisfy personal, disciplinary and task group goals by writing papers that dealt with more narrowly defined professional issues. And secondly there was a mandate in effect that had defined one of the CREAT goals as developing a set of "fact sheets" (the ten questions) for each of several contaminants. Several of the team members withdrew after these fact sheets were done. The efforts of CREAT were also split into the study of two contaminants, chromium VI and tritium.

Soil Crusts Findings

The Soil Crust RW team had only two authoring teams, one large effort to produce a paper on lichens on the Hanford Reservation, and the other a solo effort on sampling techniques in field work. The purpose of the sampling techniques paper was to discuss methodology for field studies of soil crusts. Unfortunately, the field sampling methods had already been selected, so there was little interdependency between these efforts. While there was no competition between the authoring teams, the only active participation from the Eastern Washington state members was directed to the lichen paper. Only members from Western Washington participated in the sampling techniques paper, but also contributed to the lichen paper.

Earthquake Disruption Findings

This RW had a confederation organization. By mandate, the research team was encouraged to investigate independently. A set of investigators in Southern California had collaborated on papers many times before and intended to continue that team effort with new papers. The Seattle investigator was isolated from the California team, though she had worked with them in the past. She produced a paper that was geographically focused on her locality. The Japanese team was unified by membership in (DRPI), but was split into small authoring teams of scholars in close proximity.

Discussion of Findings:

If the RW is organized as a confederation of authoring teams with little interdependency of topics, then the authoring teams will be naturally isolated. This isolation is a product of attentional economics: there is no reward in paying attention to work not related to your own efforts, and your attention will naturally be given to the paper your team is producing. This isolation can be exacerbated by geographic concentration, as the relative lack of communication barriers favors working closely with neighbors, especially people one has worked with before.

The Earthquake Disruption RW exhibited several isolating tendencies: geographic clustering, existence of previously existing authoring teams, little interdependency between topics, and perhaps language preferences. The mandate from the granting agency specified that independent research be pursued by geographically dispersed authoring teams. None of these factors contributes positively to a RW collaboration. Faced with these difficulties, this RW degenerated from an attempt at collaboration to a file-sharing web site and finally simply vanished.

This finding demonstrates the importance of an organizing model that shows how objects and processes relate to the topic of each research paper. A well-defined issue domain will support an organizing model that will show how (or if) the constituent topics are related. If there is no mandate for synthesis and collaboration backed by effective leadership, then authoring teams will tend to isolate themselves. The academic reward system is such that research papers are essentially the only professional reward, so the leadership of the RW must provide incentives for participation beyond the writing of research papers.

5.2.5.2 Research Question 2:

What were the geographic distribution effects on the RW?

Framework for Analysis:

Geographic distribution effects include not only the physical dispersion of the team, but the existence of socially bound clusters of members and isolated members of otherwise concentrated teams. Another isolating geographic influence is native language: people do prefer to work in their native tongue rather than in other languages, especially since a research paper can always be translated as a single stand-alone document. Other geographic effects include the temporal dispersion of the team: for instance, though vast distances may separate team members, they may still be in the same or nearby time zones. Separation by several time zones makes synchronous communication problematical.

Propositions:

I) RWs with strong concentrations of people who can easily communicate in person will fail.

This proposition was suggested by the media competition theory. This theory suggests that the most accessible communication modes will be preferred to those requiring more effort. Thus synchronous communication, especially face-to-face communication, will be preferred to asynchronous communication. This preference will naturally lead to a tendency for interacting with colleagues close at hand, socially and intellectually isolating remotely located colleagues[37]. Furthermore, most face-to-face and telephone

communication goes unrecorded and hence unavailable to members both remote and local.

Chromium VI Findings

The Chromium VI RW team was all located in the same city. Most of the team members were not, however located in the same office suites. The team members met frequently, and communicated by phone and informally. No indication of ill effects due to proximity was noted. There were no isolated members, since all members were drawn from a pool of people who were working on a single large grant project (CRESP). Participation in DocReviews of essays and minutes was active and successful. E-mail apparently was preferred to telephoning due to competing schedule demands. Messages were frequently shared by forwarding and multiple addressing.

Soil Crusts Findings

This RW had two members who worked at the same laboratory, and were near the lead author of an authoring team. These people were also close to the location of the field study that was the basis for the research paper. The scientific coordinator was isolated from these people, and could meet with them infrequently or individually by long-distance telephone. The scientific coordinator did actively participate in biannual fieldwork with that local group. There was a noticeable social strain in this RW, perhaps due in part to the communication problems.

Earthquake Disruption Findings

The Earthquake Disruption RW had three centers of activity with teams in Japan and Southern California and an isolated member, the scientific coordinator, in Seattle. Within the Japanese team there were four members from Kyoto, and two from other

Universities. Four of the five members from Southern California were from the University of Southern California (USC). The USC members formed a particularly tight group, having authored over thirty research papers jointly in some combination or other. The RW team then consisted of two four-strong centers and four isolated members. The extent to which this configuration contributed to the failure of the RW is unknown.

Discussion of Findings:

In the Earthquake Disruption RW most collaboration was done within the authoring groups, each isolated in its own geographic region. It is expected that authoring teams will concentrate their attentions on documents of their own. Leadership and training materials will have to remind them that all members have a responsibility to contribute to the refinement of content contributed by others.

The Soil Crust RW had a similar lack of conscription devices, though the scientific coordinator did attempt to draw members into collaboration by several weak conscription devices: an Annotated HyperBibliography, a Lexicon designed to build the Annotated HyperGlossary, and an attempt to build a Species List. The only effective conscription device was a research paper that was DocReviewed in an advanced draft. Team members expressed an interest in collaboration on the WWW, but few work objects were offered.

In order to detect problems of this nature in any RW, both isolated members and collocated groups will have to exist in the RW. This condition was not present in the Chromium VI RW. Any conclusions drawn on the limited experiences herein are conjectural and will need to be investigated as a natural experiment in the future when a RW with a widely distributed team with clusters emerges.

II) RWs with widely distributed members who live in different time zones are more likely to succeed than RWs with concentrated membership.

People who live in time zones far removed from their collaborators cannot engage in synchronous activities such as teleconferences without disrupting their daily activity cycle. This fact is likely to make a very dispersed team more inclined to accept asynchronous communication, and with that the environment of the RW.

RW Findings

The Soil Crust and Chromium VI RWs were all based within a single time zone. The Earthquake Disruption RW was a transpacific enterprise that had teams separated by 8 time zones. Recognition of this fact was perhaps a factor that caused the team to accept the concept of the RW. Unfortunately, the lack of conscription devices made participation in the RW rather pointless.

Discussion of Findings:

There is little empirical evidence to support or reject this proposition. If there is an effect, it is likely to be weak when compared to strong influences like the presence of strong leadership, member commitment, and an abundance of conscription devices.

5.2.5.3 Research Question 3:

How many people were invited to participate in the Research Web?

Framework for Analysis:

There are very fuzzy upper and lower limits to the size of effective research teams. Team size is effectively left to chance when the RW concept is adapted to existing teams rather than building the team to suit the concept for application to a specific issue domain. Here we are left with the problem of comparing the teams that were assigned by circumstance to the teams that might have been designed for the RW.

Propositions:

I) The critical mass theory holds for research webs.

Critical mass is a function of the size and organization of the research team. There are upper and lower limits to the size of an effective RW team. The upper limit is reached when there are so many scholars studying of the issue domain that scholarship is exhausted, the field becomes known territory. The lower limit, critical mass, is reached by having enough active and interdependent conscription devices to hold the interests of the entire team. In a RW, critical mass is necessary to insure a reliable flow of new content, essays, e-mail, annotation, and research paper drafts. The presence of a large body of content open to annotation (conscription devices such as models, essays, bibliographies, and glossaries, etc.) is a good base, but new content is necessary to prevent the collaboration from going stale.

Chromium VI Findings

The Chromium VI RW directly invited 18 people to participate and made the URL for the RW available to the entire CRESP team, perhaps forty people. This RW failed to recruit a full range of scholars in fields that could contribute to the understanding of the issue domain. Remediation is an important part of the study of environmental contamination. Even though there was a well-funded contingent of environmental

engineering scholars available in a CRESP task group, they did not respond to requests to join the effort; leadership was not able to persuade them to do so.

This team was very close to having critical mass. The research team was well represented in several appropriate disciplines: ecology, human health hazards, geography, statistics, and risk management. This team was comfortable with the RW technology, contributed content when requested to do so, and was not reluctant to criticize content through annotation. While the lack of participation from engineering was damaging, it may have been overcome in time, since several team members were capable of contributing essays on the topic of remediation techniques.

Soil Crusts Findings

The Soil Crust RW invited four members of the research paper authoring team, two people from a granting agency, two co-workers of the scientific coordinator, and two outside scholars. Despite a thorough briefing about the concept of Research Webs from the scientific coordinator, this small team seldom showed any inclination to participate in activities other than the authoring of a single paper. The lack of interest in any systemic studies of soil crusts ensured that there would be little likelihood of attracting new members. There was no interdisciplinary work and thus the group was not likely to attract new members. Critical mass was not approached.

This RW was started with the understanding that a recruiting effort would be required to attract critical mass. These new recruits would hopefully be drawn from scholars interested in systematic studies that would complement the existing team's interest in taxonomy and local inventories. The scientific coordinator attempted to attract these

scholars but had no success, though a few scholars expressed interest. Perhaps organizing models need to be present in order for new members to understand how their work can benefit from the RW. With no funding or colleagues with plans to produce papers, recruiting was a rather hopeless task.

Earthquake Disruption Findings

The Earthquake Disruption RW invited all twelve scholars funded by the founding research grant. Critical mass was not reached in this RW for several reasons, principally a failure of adoption of the RW as a medium of communication. The failure of adoption was in turn triggered by a lack of conscription devices. In other words, contributions were not solicited effectively.

The character of this RW was such that interactivity was subordinated to independent research. Lack of interdependence leads to a paucity of reciprocal communication. It is difficult to see how this RW could have succeeded, indeed it epitomized the "confederation" organization: a loose collection of groups working on similar, if not competing research.

Discussion of Findings:

Critical mass implies the presence of not only a sufficient number of participating scholars, but also an organization that will support collaboration. The organizational character of the RW must create a generous number of interdependent interests[38], and the conveners must have set out the terms of team membership in order to reduce the possibility of free riding and non-participation[39].

The scope of the issue domain is the key to developing interdependent units that can engage the interests of the team members. The Soil Crust RW had a scope that was perhaps proper, but did not attract enough members. There was no organizational model set up to outline potential interdependent research units within the issue domain. Without a model there was no basis to attract specialists. The Earthquake Disruption RW had no stated organizing model, and the research units were not interdependent, but rather similar. The Chromium VI Research Web had an organizing model, several interdependent research units and, very likely, enough scholars to engage those research topics. On this basis, it seems that the only RW that approached critical mass, and thus a chance to succeed, was the Chromium VI RW. It was the largest RW, and still too small.

5.2.5.4 Research Question 4:

What incentive(s) did each of the participants have to participate?

Framework for Analysis:

This research question goes straight to the heart of the reward system. For lead authors and the members of their authoring teams, clearly the incentive is to have research published. This is the well-understood academic reward system in operation. For the critics, there is the reward of showing one's peers that you do understand the issue and can contribute. Criticism is the personal expression of the mandated skepticism of science. It is also obvious that critics can be invited to participate in authoring teams if their observations are acute and well expressed.

Conveners and scientific coordinators are likely to equate success of the RW to administrative as well as scientific accomplishments. The rewards at this level are likely to be career related milestones that transcend authorship, though their close involvement assure that authorship is almost automatic. Career milestones include being selected as Principal Investigator, awarded an endowed chair, and perhaps leadership positions such as Laboratory Director.

Collaborators, those that make modest contributions to infrastructure such as model building, glossaries and bibliographies, as well as criticism, can aspire to eventual inclusion in authoring teams. Graduate students and staff assistants can earn their bread in collaboration, and can also be rewarded for their efforts by acknowledged contribution in research, a mechanism called legitimate peripheral participation[40]. Facilitators may be staff or collaborators. Their incentives may be process related, contributing to the collaboration process; and/or topic related, contributing to building knowledge in the issue domain.

Propositions:

I) In order to be successful the RW must provide rewards beyond authorship.

There is considerable overhead in a RW. The costs of the knowledge-building efforts directed to the understanding of the entire issue domain must be borne by the team. Researchers focused on the writing of a single research report cannot justify these costs. There must be some additional rewards to encourage the team to invest in model-building and collaborative criticism.

Chromium VI Findings

The Chromium VI RW though short-lived, exhibited the presence of authoring incentives. The seventeen essays provided the authors the opportunity to exercise scholarship with essays that could have provided a start of research papers in several instances. Essays on the effects of hexavalent chromium on salmon, on bioremediation of hexavalent chromium, and on the measurement of chromium contamination could have been extended into research papers.

The need to develop some rather simple research essays provided opportunities for participation without great effort. The facilitator, a graduate student, was highly motivated by the expression of interest among team members. Several graduate students and professional staff members contributed opinions and knowledge in several of the DocReviews, fulfilling not only an obligation to contribute, but also showing interest.

Soil Crusts Findings

The existence of a research paper as a conscription device certainly rewarded all the scholars on the authoring team. The scientific coordinator saw the possibility of creating not only a successful RW with multiple products, but also a career-enhancing position as host of a site that might attract new members throughout the discipline. Her research interests, soil ecology, included soil crusts. That the soil crusts were being studied as funded research associated with her position made the fit perfect. One of the stakeholders, a grant provider, was an active participant in some DocReviews. This participation was clearly offered in a collaborative spirit since he was essentially office-bound by his position.

The RW provided the opportunity for two graduates student to join the research process. One of the members was a doctoral candidate studying lichens, her dissertation topic. The facilitator not only practiced the running of the RW's site and assisting the scientific coordinator in her efforts to make the effort a success, but was also carried as a co-presenter of a paper at a conference. The Soil Crust RW provided an opportunity for the facilitator to polish some of the tools and to produce a new tool specially suited to the issue domain.

Experience with the Chromium VI RW caused one of the contributors to start the Soil Crust RW, perhaps as a career-enhancing strategy. Recently, in a career move, this person accepted an environmental consulting position. She attributes her experience with the RWs as a major factor in obtaining the position.

Earthquake Disruption Findings

The Earthquake Disruption RW offered few rewards other than those offered by the founding grant. The mounting of research papers in the public partition of the web site gave the papers wider circulation.

Discussion of Findings:

Rewards emanating from the RWs were few. Other than authorship on one research paper there were no obvious rewards save the pleasure of active participation in a collaborative enterprise. Exposure to computer-aided collaboration has been reported to be positively viewed by potential employers. There was not a great deal of effort put into development of rewards, most effort was directed toward development of the web site and intellectual content. The short lives of the RWs did not allow development of any management philosophy.

5.2.6 Conclusions

These case studies provide explanations for failures, examples of successes, and suggestions for correcting pathologies and capitalizing on successes. In most cases the RW concept was applied to preexisting teams with either inappropriate or ill-defined issue domains. The establishment of goals should precede the determination of the research team's composition. In a RW, the principal goal of a research team is always the understanding of an issue domain. Secondary goals, such as publications, will be produced as a byproduct of the search for understanding.

There are two major problems in defining the issue domain: defining a scope that is large enough to develop critical mass, and defining a scope that is small enough to ensure that the majority of authoring topics will be interdependent. The prototype, a study of migration behavior, had a scope that was far too large. The Earthquake Disruption RW had a scope that was probably appropriate, but the team members were set on parallel tracks rather than interdependent tracks. The Soil Crust RW had an issue domain but never developed an organizing model. If one had to express an issue domain, the only statement would be: anything about soil crusts, but especially those factors that contribute to our work object, a single research paper. Once the research paper reached an advanced draft, the team had no remaining goals. The Chromium VI RW had a well-defined issue domain and proper scope. It failed due to withdrawal of management support.

Critical mass was approached only in the Chromium VI RW. The issue domain was defined with sufficient accuracy to determine where the team needed to be supplemented. Had the RW lasted a few more weeks, the needed researchers would likely have been

recruited. The other RWs had small teams, but suffered from pathologies in addition to simple lack of critical mass. Based on these studies and suggestions from the literature, it seems likely that the critical mass for a RW may be as much as a couple dozen researchers.

> Proposition 1: *There are only two stable states of interactive medium usage in a community: all or nothing. Either usage will spread to all members of a community (universal access will be achieved) or no one will use the medium (for communications internal to the community), either because no one started using it or because usage fell off in the absence of reciprocity.* --- Markus 1987[41]

In her discussion of this proposition, Markus depended on a small set of natural experiments, since it was difficult to find documented evidence of participation in interactive communities at that pre-WWW time. Since then the Internet has provided ample evidence in the form of listservers and discussion forums. In our case studies, none of the communities reached a positive participative equilibrium, providing evidence that if there is a threshold (critical mass), then groups of less than two dozen or so are below it. I find no difficulty in visualizing such a threshold since people very quickly abandon an enterprise when it is failing to thrive.

Participation is a necessary attribute for success. Only one RW had adequate participation: the Chromium VI RW. Why? This team was socially integrated; most people knew each other from team meetings on several levels. The team was technologically well served and were adept users. There was an abundant supply of work object, opportunities to participate. There were 43 web pages on the web site, and most of those were available for annotation. Several members were authors of RW essays. It appears that both critical mass and frequent introduction of new content are necessary to generate adequate participation.

Geographical and time zone distribution effects were not seen in these cases. Some obvious problems centering on team dynamics dealing with cliques and isolated members remain to be investigated. Distribution did not appear to alter the technology requirements of the teams investigated.

Financial support appears to contribute to success. In a world full of interests competing for attention, money provides a simple metric for selection. Members of the successful Chromium VI RW were all supported to some degree by grant money from the CRESP project. CRESP personnel in the Soil Crust RW were weakly supported as a "management approved activity," but some members had to scramble for support or approval. All members of the Earthquake Disruption RW were supported by grants. The unsuccessful prototype RW was a purely volunteer effort.

Leadership is another quality that must be present. Though all scientific coordinators and the facilitator were enthusiastic, clear and unambiguous continuing support from the team's senior scientists was evident only in the case of the Chromium RW.

Management's role in project failure is sufficient but not necessary.

--- Charlie Hendricksen, 1983

5.3 Discussion

A synthetic work like this dissertation must eventually turn to evaluation of its products. Is the concept of Research Webs likely to survive a test in the real world? Under what circumstances is the RW effective and when not? Is the RW compatible with the

research culture and academic institutions? Is the high overhead of the RW justified by more productivity and/or higher quality?

5.3.1 The Optimal Environment for a Research Web

The fundamental assumption underlying this work is that there is a great need for methods to support large-scale long-term research. If that is true, then we can discuss what kinds of research might benefit from the concept of Research Webs; and what kinds of research will not benefit (see §5.3.2). Where does the RW fit into the existing types of research? The question of "critical mass" has arisen so often in this work that we need to discuss that issue. What disciplines are suitable for employment of Research Webs?

The Nature of the RW's Research

Theory-building research is the logical home for the RW. Its models are the expression of theory, according to the tripartite models of realism[42] and conform well to the tripartite research methodology proposed in the VNS[43]. Problem solving does not involve theory building, but evaluates proposed solutions based on existing theories, assumptions, myths, or rules. Action research can employ the RW as a theory-building activity that operates in parallel to the design of action to solve a problem and the evaluation of the implemented action[44] (see §2.2.4.6.5). More appropriately, it might be better to look at action research as a technique to be employed to investigate portions of the issue domain. Actions applied to problems are field experiments analogous to experimental scenarios submitted to simulation models.

Size

Small-scale research simply cannot afford the high overhead of the RW. The RW needs economies of scale to justify modeling, bibliographic research, glossary building and the

construction of an elaborate web site. This fact effectively eliminates the RW as an organizing method for solo and small group research.

Extraordinarily expansive issue domains, such as migration and poverty cannot employ the RW as an organization for the entire issue domain. The focus of such issues is simply too broad to be parsed into research tasks that are interdependent. Lack of interdependency reduces the potential for collaboration. Note, however, that such very large issues might contain smaller constituent issue domains that are quite suitable for treatment by the use of RWs. There are signs of this sort of organization in the MacArthur Foundation's work, where their Research Networks are all tied into on overarching objective to improve the human condition and community development[45].

The proper size for a RW's research team is an open question. In the limited body of experience, it does seem that there are definite lower limits, and that those limits are significantly higher then is usually seen in social science research. The upper limits are likely to be established more by the extent of the issue domain. The scope of the issue domain is established on the high side by the need for maintenance of interdependency, and on the low side by the presence of an adequate body of related and attractive research topics (see §3.1.1).

Geographical dispersal

The initial reasons for investigating the RW came from the circumstances that academic social science is beset with. As discussed above, isolation of specialists caused by economic necessity ensures that a critical mass of scholars can only be found by reaching beyond local sources for collaborators (see §1.1). After the investigation was entered there were mechanisms found beyond geographical dispersal that make the RW and attractive organization for research. One of the most interesting findings was that, with modern technology, because of scheduling incompatibilities, even colleagues within

shouting distance were likely to be contacted via e-mail. Once the medium of communication is asynchronous, dispersal becomes a less serious problem.

A corollary of physical dispersal is cultural difference. If the research team becomes intercultural, a number of negative factors come into play: language skills, power structures, and workday asynchrony. The pathologies introduced by all these factors are remedied to some extent in the RW. Written language skills are higher than spoken skills, power structures are blunted, and workday asynchrony becomes a minor inconvenience. The positive virtues of a culturally diverse team include a multiplicity of viewpoints.

One problem that remains uncertain is the effect of media competition. Whenever local groups, even pairs, collaborate, there is a strong tendency to revert to habits of speech and the building local tacit knowledge, thus neglecting the documentation and distribution of new knowledge. This unfortunate property may make such groups less effective collaborators in the RW. This mechanism may also increase the isolation of remotely located individuals.

Discipline

Physical sciences such as molecular biology have embraced general systems theory and from that have built some quite elaborate models of metabolism pathways[46]. The Institute for Systems Biology (ISB) in has integrated collaboration, modeling and thorough exploration of each element in the models. While the ISB has the transfer of knowledge to mankind as its goal, the presence of extremely large potential profits in industrial research may stifle free exchange of information.

Researchers in the humanities have done some collaborative work that makes good use of the current technological environment[47]. The issue domains investigated are often very

elaborate catalogs of works of art or literature[48, 49]. The need for theory building in these issue domains seems slight.

The social sciences seem to be a natural home for the RW. Theory building is necessary to account for the behavior of human behavior. The need to accommodate several disciplines simultaneously is a characteristic of large-scale social science research. The social sciences routinely have to view their interests from several points of view and inferential methods. The inexact nature of the objects, processes and measuring techniques makes critical thinking not only likely, but also essential. Criticism is a major process in content development in the RW.

5.3.2 The Research Web Compared to Conventional Research Teams

The question of the value of Research Webs must be answered here. To provide a framework for evaluation of the Research Web concept, this section assumes the existence of two equally staffed research efforts, both long-term and large-scale. One of the efforts is assumed to be a conventional research team, perhaps a university institute, the other a team using the Research Web concept and tools.

Evaluation factors that remain similar for both teams:

Salaries and support requirements for scholars, administrators, research assistants, and a copy editor should be identical. The technical support environment (hardware) should be identical. In the academic environment WWW servers are part of the university infrastructure. Software costs are only very slightly higher for the RW, and any higher cost is miniscule compared to salaries. Travel costs may be higher in an RW because it is more likely to be dispersed; but, if the conventional team is dispersed as well, then its travel expenses will likely be greater than the RW team's expenses, because the RW has much better communication channels and therefore less need to travel.

Factors involving personnel cost:

Facilitator

The active presence of a facilitator is absolutely essential. Every member of the team must know that there is someone there to take on all the mundane tasks that the content providers are best relieved of. As Marwell and Oliver say about collective action[50], "The most important determinants of collective action in our models are the interest and resource (or contribution size) levels relative to the cost of contributing." The duties of the facilitator are simply to assume the technical workload introduced by the necessity to support the web site, models, and document formatting. The facilitator reduces the cost of contribution to the team every time one of the many gritty technical chores is taken off a contributor's plate.

From a management point of view the facilitator is not only a cost, but also a bottleneck in some aspects of the information flow[51]. In the RW, most communication is not mediated or facilitated, but some tools are designed under the assumption of intermediation. Access to those tools is by password, and should any team member care to operate the software, then the password can be shared. Experience shows, however, that most members are uncomfortable with the need to learn software and also are reluctant to spend time on such tasks.

Modeler

The services of a modeler may be required, probably on a part-time basis, especially if a simulation model is being produced. While the researchers will certainly become comfortable with the models from examples, there are nuances of data modeling that require more specialized knowledge. For the descriptive, auxiliary, and explanatory models, any moderately skilled programmer/technician should possess the necessary skills. The facilitator or a research assistant may assume these duties. In a mature RW, the services of a more highly trained modeler will be of great value. A skilled modeler can bring a rigor and intellectual clarity to the process that will increase the value of the models.

Training for Collaboration

It should not be assumed that members and staff have any experience with collaboration in large groups. As discussed in the preceding chapters, the experiences of most researchers are likely to have been more competitive than cooperative and limited to occasional small-scale shared writing (see 2.2.5.4). Training that introduces the team to the RW concept and the value and costs of collaboration must be undertaken early in the life of the RW and should be repeated as new members come on board. The cost of this training must be balanced against the value of acquired skills. I think it is likely that a couple of hours of training will be an unquestioned bargain.

Web Site content creation and maintenance

This expense is unique to the RW, although the conventional research team might have a "brochure" web site to publicize its work. A brochure site is equivalent to a token public partition (see §3.5.1) of the RW. The RW's team working area (see §3.5.2) and guest partition (see §3.5.1) have no equivalent in the conventional research effort. Many of the RW's software tools have been built with the intent of reducing the work required to create and maintain content. Much of the content of the RW's web site is actually collected and installed automatically by software. The costs and benefits of this content are discussed below under the "Artifacts Produced" headings.

Collaborative review of documents

In the Research Web team members are expected to review documents as the team produces them (see §2.2.4.6.7). This review process informs the reader and allows the reader to contribute insights that the authoring team may have overlooked, especially links to other research, not to mention the occasional correction of error or the offer of a reference to support a point. A trained collaborator will come to understand the value of critical reviews when his or her work comes under review. In conventional research, the

obligation to review the work of those outside a single authoring team is likely to be minimal.

Artifacts Produced

Models

Models in conventional research are, if present, usually sentential or tacit and deal with only the topic of a single research paper. A good model provides a description that is far more complete than the usual abridged and partially tacit narrative description found in conventional reports.

> *"A fairly common strategy in group research is to conduct studies to explore some interesting phenomenon or technique. ... Each of these has produced a 'minitheory' of the phenomenon under study, and some of these are ingenuous and useful. However for the most part, these efforts do not attempt to tie into a larger theory and therefore remain isolated findings."*
>
> --- Poole 1990[52]

In the RW, the "minitheories" are expressed, connected, annotated and thoroughly documented. More importantly, they are strongly connected to other minitheories in the RW's issue domain. The question that needs to be answered is, "Is the modeling effort worth the expense?" We discuss below each of the several models that exist in a mature RW.

The Descriptive Model

In conventional research, the objects and processes that apply to the topic of a research paper, usually an isolated phenomenon, are described by reference and narrative description in the research report. No attempt is made to describe attributes of the objects that are not essential to the report. These ignored objects include superclass objects (embedding objects) and objects that might have an unknown or presumed infinitesimal

effect on the phenomenon. Process models concern themselves only with elements essential to the argument of the paper.

In the RW, every object within the issue domain that is mentioned in any document is described in the Descriptive Model (see §2.1.2). Every attribute of objects that suggest themselves to the members is listed; and attributes found of interest are elaborated in appropriate detail including operationalization methods. Every object's attributes are open to annotation by the research team. Since several authoring teams share the same models, attributes that are of no interest to one team may be essential to another. Furthermore, some attributes may constrain others.

Processes models are descriptions of observed behavior of the objects. Like obscure attributes of objects, processes may have "side effects" that are of little interest to the observers of the phenomenon being investigated for a single paper. In a RW, one team's side effects may be another team's phenomenon of interest.

The Auxiliary Model

In conventional research, the auxiliary model is generally restricted to narrative description of the operationalization of variables used in experiments and analysis of data. In the RW, all attributes of objects used in experiments, data analysis, or the simulation model are formally operationalized. The investigation is recorded by adding an extension to the Descriptive Model. The extension will include measurement criteria for the attribute: data type of value, precision, range, and a description (see Table II §3.2.4.2.1). In the auxiliary model several options for operationalization may be offered. Choosing the appropriate operationalization is the duty of theoreticians and experimenters. The auxiliary model will be open to critical annotation by the team, just as any document in the RW. The operationalization of the attribute is a directive to the programmer managing the simulation model.

The Explanatory Model

Conventional research expresses its explanatory model as a narrative discussion leading to the development of hypotheses. The hypotheses are the basis for experimentation or argumentation. In the RW, the explanatory model is likely to be somewhat more formal, but incorporates all elements of the explanatory model presented in conventional research. The hypotheses investigated in conventional research are related primarily to the topic of the single research report. The great difference is in the scope of the model. The explanatory model may be abridged for inclusion in research reports. In the RW, the team contains many authoring teams, each intimately familiar with the phenomenon under investigation in their own paper. The RW's explanatory model (see §2.1.3) will show how each phenomenon is related to another. Examination of the relationships is likely to modify the conceptualizations of the phenomena. Certainly additional hypotheses will be suggested.

The Simulation Model

Simulation models are infrequently produced in conventional research in social sciences. Their value is unquestioned in the physical and natural sciences. The reasons for the rarity of simulation in the social sciences are many and varied, but they center on modeling's frequent failure to produce reliable results. The RW can mitigate many of the likely elements of failure. First, the objects and their operationalization are better described and thoroughly discussed; second, the process models of the explanatory model that form the basis for the dynamics of the simulation are more accurate than the isolated "minitheories" of conventional research; and finally, the entire team will be able to "exercise" the model and thus expose potential problems. An argument could be made that high quality simulations might open the phenomena of any issue domain to experimental social science, just as has happened in some domains in economics.

The Context Model

The context model is the most abstract of all the proposed models. It describes the scope of the issue domain (see §3.1.1). In conventional research the contextual model is likely to be expressed only in the charter of the sponsoring organization. This expression is likely to be carefully crafted for political and organizational purposes. The political aspects mentioned are designed to demonstrate the moral grounding of the sponsoring organization. The description is likely to be purposely vague in order not to alienate too many sources of grants.

The context model in the RW has definite purposes. Precision in establishment of boundaries is important for two reasons: if the issue domain is too small, then the issue domain may become unproductive -- all potential research topics may be used up; if the domain is too large, then the authoring teams may not be interdependent --- the research team will be fragmented to the extent that critical mass is not maintained. The representation of the context model is still unexplored territory, but is likely to benefit from cognitive mapping[53,54] and concept mapping. The model must be able to establish

not only the necessarily fuzzy boundaries of the issue domain, but also the adjacencies to other issues. The RW's research team will be in a position to contribute knowledge or conjectures about those boundaries and adjacencies.

Intellectual content

In conventional research, intellectual content takes the form of partial drafts, notes, outlines and other ephemera that eventually lead to a research paper. This mass of expressed knowledge is usually stored in the author's files, and is thus inaccessible. After publication of the research paper, much of this information is purged from the files. Even if the files were opened to the authoring team, they would not be searchable, and would thus present the reader with the necessity of wading through every scrap of information in order to find what is of interest.

In the RW, intellectual content is contained in the documents in the RW repository (see §3.2.2). Documents include essays, models, e-mail, drafts of papers, outlines, images, annotations, and others. Essays are a formal genre, the Research Web Essay (see §3.4), and are accessible to and annotatable by the entire team. They are organized in a hierarchy of web pages, linked hypertextually from graphic models, and subject to full text search; and the essays are annotatable. Notes may be expressed in e-mail messages that are in the team's searchable e-mail archives. Annotations are documents as well and can be found through full text searching as well as through DocReview. Outlines are a fundamental scholarly tool, a model of a document. In the RW outlines cannot only be annotated, but can form the backbone for preliminary drafts built by annotation with DocReview. Models are expressed in graphic and textual documents and are both searchable and annotatable.

Publications

Conventional research is focused on the production of research reports, often just the smallest publishable unit. Of course the strategy of least publishable unit could be followed in the RW as well. The RW views research reports as derivative, not as the sole objective. The RW can serve not only to produce research reports, but monographs and books as well. Even criticisms of reports or books in the literature can be generated by the team, leading to publication as brevia, reviews or letters to the editor. Such criticism of external materials can be archived and linked to documents in the RW repository.

It is extremely difficult to argue that the RW's research team would generate fewer publications than a conventional research team. While it is true that the effort expended in model building does not lead directly to publication, it is also true that the shared models may reduce the total expenditure of effort. The marginal cost of refining a model to include information generated in the research of a phenomenon is very small compared to the cost of reproducing the models in each research report investigation. I believe that contribution to models and the study of models will familiarize all the researchers with the issue domain to such an extent that many more hypotheses will be generated. These hypotheses are each a potential research report topic.

The purpose of a model is to rub the researcher's nose in the problem.

----paraphrasing a comment by Professor Dick Hamlet[55]

Bibliographies

In conventional research, as well as in the RW, an enormous quantity of information is gathered from the literature. The lists of references from research reports are usually the only permanent bibliographies produced in conventional research, references for literature investigated but not cited is lost. In the RW, the research team has several advantages that scholars in the conventional environment do not have. Possession of a corporate body of literature indexed and searchable by several methods will save

considerable duplication of literature search effort. In the well-disciplined RW, every piece of literature examined by the team members can be incorporated into the Annotated HyperBibliography (see §4.4). The RW's team has the ability to read abstracts and full text at a click along with the opinions of colleagues. If a document is seen to be less than useful then the member who reported can annotate the entry with a warning to avoid it. With the assistance of the facilitator, bibliographic entries can be tagged so special purpose bibliographies can be produced. Hypertext links can be made to the bibliography from any document in the RW, even e-mail. The Annotated HyperBibliography can be searched by author, title, keyword, or text in the abstract.

Glossaries

Glossaries are seldom included in research reports, even though readers from disciplines outside the discipline of the authors may not be familiar with the intended specialized meanings of terms. The appropriation of widely used words for more narrowly defined disciplinary meanings is both necessary and widespread. In the RW, the team members have an Annotated HyperGlossary (see §4.5) at hand to investigate the language of the dialog. Should a scholar use a word in a technical sense, that word can be linked to the proper definition, in interdisciplinary teams often one of several. Should a RW team member choose to add a gloss or alternative meaning to the HyperGlossary, the facilitator can do that. The Annotated HyperGlossary provides the RW team with a forum to discuss the terms through annotation

.

Services Offered

Facilitated document review

Conventional scholarship does not usually have the benefit of shared criticism; and review is often limited to electronic file swapping of drafts between the lead author and the authoring team. In the RW, the use of DocReview (see §4.3) provides the entire team the opportunity not only to criticize most documents, but also to discuss the commentary

of others in a focused team-wide forum. Through the guest partition (see §3.5.1) selected scholars outside the team may be invited to review documents that would benefit from a review by qualified specialists.

Proposal production

The problem of controlling the costs of proposal production is widely recognized. Indeed the cost of producing a proposal for an Institute-level grant is very expensive and is cited as a barrier to commissioning such organizations[56]. Software is becoming available to manage the paperwork of preparing a proposal, but I am aware of no services that attempt to provide intellectual guidance to the proposal team. The mature RW has a very valuable repository of information that is available to support the preparation of the intellectual content of the proposal. Since any proposals generated by the research team must be related to the issue domain, there exists a set of points of contact with existing research and the models of the issue domain. The proposal can actually point to the connection of the proposed work to existing work within the organization as well as to the outside literature.

Technology training

In conventional research teams, any training of team members is provided by the infrastructure of their work environment. In the RW, the facilitator is directly tasked to provide any technical training that a team member may request. The RW is a more technical environment, but most tools appropriated should be web-based, thus the training required is minimal --- every page is a web page. Becoming familiar with modeling is quite another issue. While most representations of models are straightforward and can easily be used as templates, a real understanding of models may require some additional training.

Wider distribution of research documents

Conventional research generally publishes only documents that meet the standard of minimum publishable unit. The RW can publish any document regardless of its scope. The advantages of wider distribution of scholarly documents are now universally recognized. The RW distributes not only the documents themselves, but also links to the participants' home pages, thus informing others of their research interests. Many of the RW's documents may be available for WWW distribution from the public partition (see §3.5.1). Always requesting all the commercial search engines to index them aggressively publicizes those documents. The important research reports may be presented in several formats: in PDF format, for direct imaging of the document; in HTML for presenting the document in hyperdocument format, thus offering the reader the advantages of sidebars, the hyperbibliography and hyperglossary; and finally, in DocReview format, allowing the wider world to annotate the document. Many documents will not be published in journals; and others will be published in journals that may not be available to many scholars.

Central e-mail repository

In conventional research, e-mail is not universally archived. In the RW, unless e-mail is private, all email is archived and is searchable. This brings a new source of intellectual content to the research team; all that is needed to recall information are keywords or an approximate date.

5.3.3 The Tools of the Research Web

Finally we dispense with the technology. The Internet and World Wide Web have provided a home for the application programming that will make the Research Web possible. There is no argument that can object to the efficacy of the large body of software that is available to facilitate any kind of research on any topic. If the software isn’t there, it can be built. Do the tools proposed for the RW really serve the research

team? The core tools proposed for the RW: DocReview, RW Essays, Annotated HyperBibliography and Annotated HyperGlossary, have been placed in service and were found to be both useful and accepted by researchers. While only DocReview received enough use to collect empirical data and support research questions (see §5.1.1), no negative opinions were received on the other tools (aside from an unusual total rejection of the technology by two senior researchers). Both the Annotated HyperBibliography and the RW essays have received praise from members of the academic community. The real question here is the efficacy of the Research Web concept.

Notes to Chapter 5

---

[1] Henri 1991, 126

[2] Meyers, *et.al*. 1991, 53

[3] Sheard 2000

[4] Bales 1950, 9

[5] Meyers, Seibold and Brashers 1991

[6] Bowker and Star 1999, 10

[7] Cohen 1960, 38

[8] Bales 1950, 177-195

[9] Meyers *et.al*. 1991, 54

[10] Bales 1950, Chapt. 2

[11] Bales 1950, 93-99

[12] McGrath and Berdahl 1998

[13] Bales 1955

[14] Geri Gay *et.al*.1999

[15] Eisenhart and Borko 1993

[16] Toulmin, Rieke and Janik 1979

[17] Meyers, Seibold and Brashers 1991, 50

[18] Meyers *et.al*., 45

[19] Meyers, Seibold and Brashers 1991, 54-55

[20] Meyers *et.al*. 1991, 56

[21] McGrath and Berdahl 1998

[22] Bales 1950, 101

[23] Nyerges *et.al*. 1998, 141

[24] Nyerges *et.al*. 1998

[25] Cohen 1960

[26] Landis & Koch 1977

[27] Perreault and Leigh 1989

[28] Cohen 1960

[29] Landis and Koch 1977, 165

[30] Perreault and Leigh 1989

[31] Zinsser 1980, 111

[32] Strunk and White 1979, 15

[33] Nielsen 2000, 110 *et.seq*.

[34] Peterson, 1977

[35] Hendricksen, 1994

[36] Link *et.al* 1999

[37] schraefel *et.al*. 2000, §2.2

[38] Markus 1993, 503

[39] Marwell and Oliver 1993, 36

[40] Lave and Wenger, 1991

[41] Markus 1987, 500

[42] Aronson, Harré and Way 1995

[43] Brinberg and McGrath 1985

[44] McKay and Marshall 2001

[45] MacArthur Foundation 2000

[46] Ideker *et.al.* 2001

[47] Ruhleder 1995

[48] Eaves 1997

[49] Landow 2002

[50] Marwell and Oliver 1993

[51] Professor Kirsten Foot 2001, personal communication

[52] Poole 1990, 239

[53] Kosko 1986

[54] Wellman 1994

[55] from a lecture by Professor Hamlet of Oregon Graduate Center, 1982. The original line was "The function of a software engineering methodology is to rub the programmer's nose in the problem."

[56] Task Force on Enhancing the Research Environment 2001, 11