Document Review Meets Social Software and the Learning Sciences

David G. Lebow, Ph.D.

CEO/President, HyLighter, Inc.; 1918 Wahalaw Ct.
Tallahassee, FL 32301

david@hylighter.com

Keywords: collaborative document review, hylighter, social annotation system, social software, threaded discussions.

Abstract

Collaborative review of documents is an integral process in many fields and enterprises. Yet, most knowledge workers are dissatisfied with current practices and products, especially when several or more contributors are involved. Today, a new approach to document review is urgently required to meet the growing needs of the global economy and to improve the outcomes of document-centered group work. Although many products available today have provisions for shared annotation, co-editing, and co-authoring of content, a satisfactory solution remains elusive for a variety of reasons. A product developed by the author and others offers a unique solution to the problem of document review by enabling several or more contributors to carry on threaded discussions tied to specific sections of a page. Such capabilities enhance the document as a medium for negotiating meaning and, potentially, provide a practical approach to harnessing the cognitive and social interaction potential of groups.

1 Introduction

The collaborative review of documents is an integral process in many enterprises and workflows. In doing market research over the past two years, we consistently found that knowledge workers in a wide variety of fields and departments are almost universally dissatisfied with current collaborative document review practices and products. Our interviews included, among others, chief information officers, chief financial officers, product managers, proposal managers, technical writers, product development engineers, in-house legal counsel, in-house marketing managers, coordinators of pharmaceutical trials, in-house social software evangelists, university administrators, faculty, and students.

More formal evidence supporting our conclusion comes from a recent study of 394 business leaders worldwide. The study was designed to gauge the extent to which the private and public sectors are encouraging, managing, and measuring collaboration. The main conclusion was that companies are facing a new imperative to form collaborative relationships. However, despite the promise of information technology and social media, dissatisfaction with online collaboration tools is widespread (Cisco Systems, 2007). In other words, even after 30 years of groupware, collaborative software, and, now, Web 2.0 social software, online collaboration and, more specifically, collaborative review of documents, remains a painful process.

Today, a new approach to document review is urgently required. In the global economy, collaboration involving groups of geographically dispersed knowledge workers is increasingly common, and the problems that we face are more complex and interdisciplinary. Finding innovative solutions to big problems requires more input from more people across multiple areas of expertise. Although many products available today have provisions for shared annotation, co-editing, and co-authoring of content, a satisfactory solution remains elusive for a variety of reasons. A product developed by the author and others offers a unique solution to the problem of document review by enabling several or more contributors to carry on threaded discussions tied to specific sections of a page. Such capabilities enhance the document as a medium for negotiating meaning and, potentially, provide a practical approach to harnessing the cognitive and social interaction potential of groups.

2 State-of-the-Existing Art for Document Review

Imagine that you have responsibility for a proposal, research report, strategic plan, or whatever form a document might take within your organization and workflow process. In order to complete some task related to the document, you

need input from several or more people. Suppose the task not only requires the group to suggest revisions to the document but, also, to analyze and evaluate the content, build consensus, and generate new knowledge or solutions. Such tasks and related purposes are not well-supported by existing products.

Most likely, you have tried one or more of the following methods when you have required input on a manuscript or draft document:

Serial Review. You create a list of reviewers and email a Microsoft (MS) Word document or PDF as an attachment to the first person on your list. The first person recommends changes using MS Track Changes, Acrobat Sticky Notes, or other mechanism and sends the marked-up document to the next person on the list. If a specific reviewer neglects the document for a couple of days (which is not uncommon), subsequent reviewers may be unavailable for a few more days, whereas they were available earlier in the review cycle. Eventually, you receive the document from the last person on the list with the accumulated markup of the group. If the review group is larger than two or three people and has made many comments and proposed changes, the document may be overwritten with markup and frustrating to decipher.

Parallel Review. You email a document to a number of people. Each person adds changes using MS Track Changes, Sticky Notes, or some other mechanism. You bring all the marked up versions together and compare each version to a master copy and make changes accordingly. Parallel review is inherently inefficient as reviewers cannot see the input that other reviewers have made, because those comments are sent directly to the author. Essentially, each reviewer responds to an outdated version of the document. This approach, as in serial review, leaves contributors isolated from each other and the owner of the document with the task of reconciling recommended changes across multiple versions. Moreover, in both serial and parallel review, redundant versions of the document remain stored on the reviewer's hard drive and in email systems, causing confusion at a later date as to which version is authoritative. Of note, Word 2007 has a new "Document Compare" feature that facilitates this process, and several products offer major improvements such as PleaseReview[1] and Textflow.[2] However, these approaches represent incremental improvements to existing practices rather than innovative solutions.

Shared Workspace. You post a document on a team space server, such as MS Office SharePoint Server (i.e., a Microsoft product with over 100 million installations worldwide) and invite reviewers. One reviewer at a time downloads the document, adds markup using MS Track Changes, and returns the document to the team space. As the document passes from one reviewer to the next, the program displays the accumulating input of contributors. Although

1 PleaseReview. See http://www.pleasereview.com/

2 Textflow. See http://www.textflow.com/

this process allows reviewers to see each other's recommendations, a checked out copy of a document may sit on a reviewer's desktop for an extended period and the opportunity for reviewers to interact is minimal. Also, as in serial review, the document becomes increasingly overwritten with markup as the number of contributors increases.

Collaborative Authoring: You and your team author a document on a website which allows users to add and edit content collectively. This is supported by wikis and online editors such as Google Docs[3] and Zoho Writer[4]. Wikis can be quite effective for such activities as sharing and cataloging information useful to a broader community or co-creating documentation for a project. Arguably, wikis make collaborative editing of a document easy, and the learning curve is, generally, only a few minutes. Similarly, online editors allow multiple users to work on a document simultaneously. However, if you are looking for input and discussion among reviewers before changes are made, wikis and online editors are not adequate solutions. Moreover, wikis, and other social software are designed for creating and revising web pages and not for working with "documents" such as MS Word documents, Excel spreadsheets, PowerPoint presentations, PDF documents, and JPEG images.

Personal Publishing: You post information on a website and readers may add comments, usually at the bottom of the document. This is supported by blog software which has value as a communication medium but is limited as an approach to collaborative review. What is lacking is a way to tie the discussion to specific sections of the text. The same is true of discussion boards and online chat which are sometimes used in conjunction with wikis and online editors as a discussion medium.

Synchronous Conferencing: Using a web conferencing product, you post a document on the web and schedule a time for team members to get together with voice communication and screen sharing. Among other issues, this defeats the advantage of using asynchronous technology to allow people to work together without the requirement to be together at the same time.

In sum, deficiencies of various document review practices and supporting technology in common use restrict the behaviors and interactions of participants and reduce the potential value of such activities. In other words, the design and implementation of the developers restricts the behavioral options and interactions of users (Flores, Winograd, 1986). Arguably, MS Track Changes as the predominant document review product has shaped the prevailing concept of document review practices to reflect its underlying design rationale of "co-editing a document with a few people." A broader concept of document review is to engage readers as writers in a collaborative and generative conversation tied

3 Google Docs. See http://en.wikipedia.org/wiki/Google_Docs_&_Spreadsheets

4 Zoho Writer. See http://writer.zoho.com/jsp/home.jsp?serviceurl=%2Findex.do

to specific sections of the document. The underlying design rationale for this alternative perspective recognizes values which are consistent with advances in the learning sciences and affordances of Web 2.0 technology. These include, among others, critically evaluate ideas, expand on what we know, trigger new ideas in others, identify areas of agreement and disagreement, build consensus, make decisions, enhance learning, improve performance, promote innovation, and build and sustain communities.

3 Historical Antecedents for a Document Review Solution

Over 5000 years ago, Sumerians inscribed wedge marks into wet clay tablets to make lists of supplies and financial accounts. Since then, people have made documents in many different ways and for an ever-increasing variety of purposes. In the Middle Ages, a type of low-tech shared annotation process was widely practiced by scholars who used the margins and spaces between lines of manuscripts to engage in dialogue with other readers. The same physical copy of a manuscript was passed around a community, and readers used the margins to correct errors, debate interpretations, and learn from the annotations left behind by previous readers. In some instances, the annotations of readers would completely engulf the primary text on all four sides of a page. When scribes made new copies, selected annotations were customarily retained (Wahlstrom, Scruton, 1997; Wolfe, 2001).

With the arrival of the printing press and movable type in the fifteenth century, the printed word became the primary means for the spread of ideas and ideologies. As Gutenberg's invention revolutionized the spread of information, shared annotation practices largely faded away. The role of reader as co-author and member of a community engaged in a collaborative search for meaning generally changed to a largely private activity. At the same time, annotation practices became more personal, idiosyncratic, and ubiquitous as demonstrated by Marshall's (1998) analysis of used textbooks from a college bookstore. In sum, the document as a medium for knowledge-production distributed between readers and writers was lost (Wolfe, 2008).

With the arrival of the computer and networking technologies, the storehouse of human knowledge began to expand rapidly and move from largely printed matter toward largely text-based and image-based digital archives. At the same time, shared annotation practices re-emerged as various forms of online annotation systems spread across the digital landscape. Today, thousands of systems exist that allow users to annotate web-based or other data via the Web or other Internet protocol for a wide variety of purposes (Shabajee, Reynolds, 2003).

Despite the proliferation of annotation systems, two problems with annotation interfaces have stood in the way of a viable solution to the problem of document

review. Real estate (i.e., the margins of the page) available for placement of commentary and extended threaded discussions among reviewers is limited. As more participants add more annotations, the challenge for the designer is how to keep commentary tied to targeted text or objects without creating separation between the two and adding cognitive effort for the user. At the same time, as more participants add highlighting and other markup to the primary text or image, the accumulating markup (i.e., highlighting, strikethroughs, and other marks - often with different colors to represent different contributors), may overwrite the document and create a confusing mass of metadata.

4 A Solution to the Document Review Problem

HyLighter is a unique browser-based document review solution. It is a software application developed by a Company having the same name (www.hylighter.com).[5] HyLighter enables virtually any number of users to carry on conversations in the margins of a text (or image) tied to specific idea units or segments of the text. In addition, HyLighter provides a variety of methods for organizing, analyzing, and editing input from the group. The design embodies certain conjectures about learning and social context drawn from the existing research and theory base of the learning sciences including, for example, social constructivism (Lave, Wenger, 1991), the theory of expert performance (Ericsson *et al*, 1993) and knowledge-building communities (Scardamalia *et al*, 1989).

HyLighter works with essentially the same data or information as MS Track Changes, online editors, and wikis. The difference is in how the information is displayed. As discussed previously, MS Track Changes is an effective tool for co-editing a document with two to three people. It is not designed for groups to have collaborative conversations. Wikis and online editors make it easy for groups to collaboratively create and revise documents but hard for members to share their thoughts before changes occur. Blogs and discussion forums allow threaded discussions for groups, but the discussions are not tied to specific sections of a page or primary source document.

What makes HyLighter different are (a) its capacity to align commentary with related sections of a page, despite the limited real estate available in the margins, and (b) its mechanism for "mapping" the intellectual travels of a group through a document (e.g., Word, HTML, Excel, PowerPoint, PDF, JPEG, and GIF) by adding a layer of color-coded highlighting on top of each selected section linked to comments in the margins. This technique enables several or more people (e.g., a proposal team, a research and development team, a distance learning class of 30 students or more) to engage in collaborative conversations tied to specific sections of a document without overwriting the primary text or cluttering the margin.

5 The development is lead by the author, David G. Lebow, and a dedicated team of programmers led by Kamal Muthuswamy.

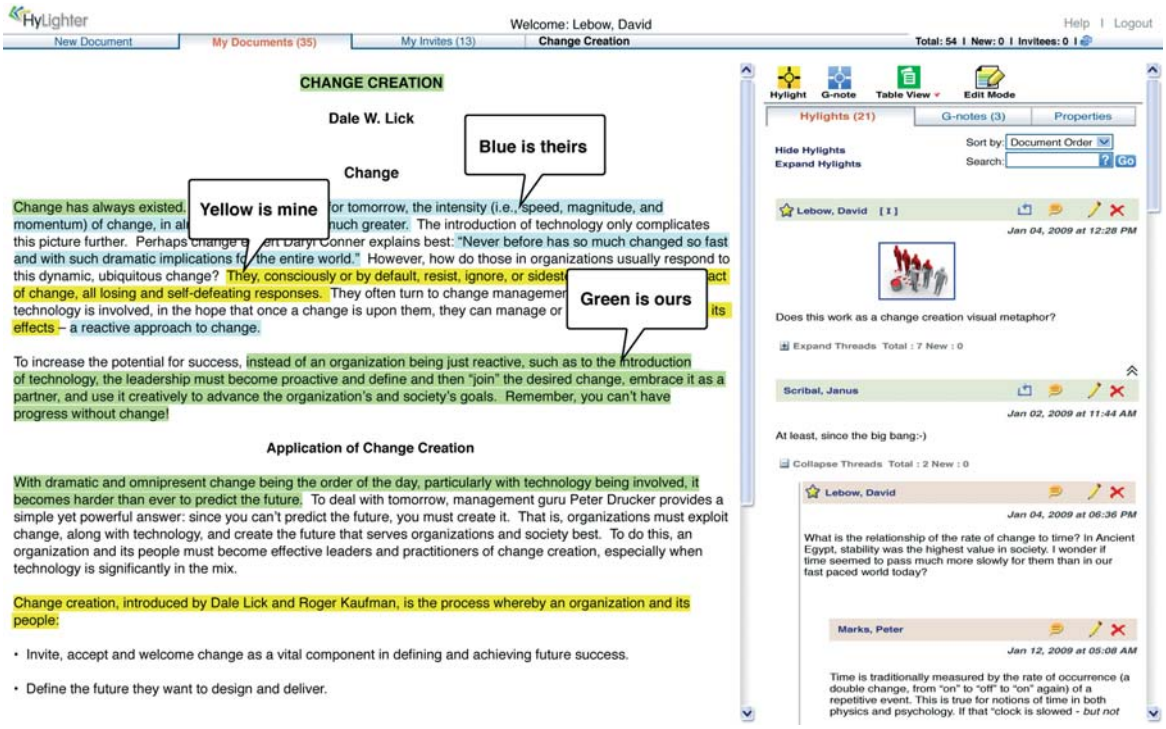

Figure 1 The HyLighter screen. The screen is split into two panels. The left panel holds the document text with various color-coded sections. The right panel shows 1) at the top, the HyLighter menus and commands; 2) below, comments by the various contributors, linked to highlighted sections of the text. Each comment is associated with a text editor bar. Threaded comments are shown indented.

For an analogy, consider the Waag Society's Amsterdam RealTime project (Scanlon, 2003). The project tracked 75 volunteers using global positioning technology for 40 days as they walked, cycled, bussed, and drove around Amsterdam. The data is displayed on maps in such a way that the more traveled a route, the brighter it glows. Similarly, HyLighter is able to create a cumulative map of multiple readers' intellectual travels through a document, and, additionally, to single out an individual's journey and compare it with the whole or with the paths of the most experienced travelers. This is accomplished by displaying sections of a document highlighted by a logged in contributor in relationship to sections highlighted by all or selected contributors using a simple color-coding mechanism. From a technical perspective, HyLighter converts supported file types imported to the program into an internal XML format, stores information added by contributors in a relational database, and overlays color-coding on selected pages as requested.

Figure 1 shows HyLighter's unique approach to mapping the distribution of highlighting (i.e., areas emphasized and commented on by readers) through a document. An area highlighted by you (the logged in user) but not by anyone else appears in yellow; areas not highlighted by you, but marked by one or more contributors, appear in shades of blue (the darker the shade, the more overlapping interest for that fragment); and areas highlighted by you and others appear in various shades of green. The margin on the right shows comments linked to the highlighted areas. Threaded comments appear indented under the original comments.

As participants engage in discussions, HyLighter provides a variety of "views" for working with the group input including (a) a table showing highlighted excerpts and associated comments with various sort options (i.e., by document sequence, username, and date modified) and search options (e.g., by key terms, recommended changes, and tags) and (b) a Word-like, online html editor for revising the document with reference to input from the group and with provisions for exporting back to MS Word or back to HyLighter for further review.

By supporting generative discussions among contributors linked to specific sections of a text or image, HyLighter increases social interaction and the knowledge-building capabilities of groups. HyLighter not only represents an alternative to MS Track Changes for co-editing documents but, also, embodies an expanded definition of document review. This richer definition of document review encompasses the sharing of multiple perspectives in order to identify areas of agreement and disagreement, expand what we know as a community, make important decisions and accomplish other purposes critical to learning and performance of individuals and the enterprise as a whole.

In sum, HyLighter tightly binds online asynchronous or synchronous discussions to specific sections of a text or image and helps to extract value from the collective thinking and interactions of contributors. Much as in the tradition of scholars during the Middle Ages, HyLighter blurs the line between authors and readers and brings rich social interaction and multiple perspectives to formerly static environments. In other words, HyLighter makes the reader's thinking that is ordinarily hidden, become "transparent" and easily accessible for sharing with others and self-reflection.

5 Conclusion

Internet "juggernauts," Google and Microsoft, see the future of information technology in an ever-expanding digital storehouse of human artifacts linked together in one searchable information universe. In this vision, sprawling server farms will provide anytime, anyplace access to virtually everything ever writ-

ten or recorded. As this vision moves toward becoming a reality, individuals and entities who learn to leverage this information storehouse will be the big winners. From this perspective, innovations in the use of social software and related technology will not only give individuals and teams smarter, faster ways to create new value but, also, change the cognitive architecture (i.e., how information is organized inside the mind) of people who engage in these practices. The result will be a competitive advantage for those who employ social software and cognitive technology across the enterprise to build capacity of the community. In sum, HyLighter is an emerging set of tools and methods that support various document review practices to harness the cognitive and social-interaction potential of knowledge-based social networks and accelerate learning, creativity, and improvements in performance of members. At its core, HyLighter builds on the principle that our own views grow and are enhanced by remaining open to the views of others.

BIBLIOGRAPHY

Cisco Systems (2007), URL: http://tools.cisco.com/dlls/tln/page/research/detail/rs/2007/collaboration , *Collaboration: Transforming the Way Business Works*, URL: http://my.elis.org/mod/resource/view.php?id=6705 (accessed on 1st january 2009).

Ericsson K. A., Krampe R. Th., Tesch-Römer C. (1993), *The role of deliberate practice in the acquisition of expert performance*, Psychological Review, 100: 363-406.

Flores F., Winograd T. (1986), *Understanding Computers and Cognition: A New Foundation for Design*, Norwood, NJ, Ablex.

Lave J., Wenger E. (1991), *Situated learning: Legitimate peripheral participation*, Cambridge, UK, Cambridge University Press.

Marshall C. C. (1998), *Toward an ecology of hypertext annotation* on Proceedings of ACM Hypertext '98, Pittsburgh, PA, June 20-24, 40-49.

Microsoft Office SharePoint, URL: http://www.microsoft.com/sharepoint/prodinfo/what.mspx (accessed on 1st january 2009).

Scanlon J. S. (2003), *Emerging Traffic*, Wired, 35.

Scardamalia M., Bereiter C., McLean R. S., Swallow J., Woodruff E. (1989), *Computer supported intentional learning environments*, Journal of Educational Computing Research, 5, 51-68.

Shabajee P., Reynolds D. (2003), *What is annotation? A short review of annotation and annotation systems*, URL: http://www.ilrt.bris.ac.uk/publications/researchreport/rr1053/report_html?ilrtyear=2003 (accessed on 1st january 2009).

Wahlstrom B., Scruton C. (1997), *Constructing texts/understanding texts: Lessons from antiquity and the Middle Ages*, Computers and Composition, 14(3), 311-328.

Wolfe J. L. (2001), *Pedagogical uses of annotation and annotation technologies*,

Dissertatio Abstracts International (UMI No 3008469).
Wolfe J. L. (2008), *Annotations and the collaborative digital library: Effects of an aligned annotation interface on student argumentation and reading strategies*, International Journal of Computer-Supported Collaborative Learning, *3*(2), 141-164.

AUTHOR

David Lebow is the CEO/President and co-founder of HyLighter, Inc. He is also the inventor of HyLighter. David received his Ph.D. in Instructional Systems Design from Florida State University in 1995. His research efforts have focused on design and development of computer-supported learning communities, and he has taught courses on alternative methods of teaching and learning.