Plot-polling: Collaborative Knowledge Visualization for Online Discussions

Alex Ivanov, Thomas Erickson, Dianne Cyr
School of Interactive Arts and Technology, Simon Fraser University,
IBM T. J. Watson Research Center,
Faculty of Business, Simon Fraser University
alexi@sfu.ca, snowfall@acm.org, cyr@sfu.ca

Abstract

Online communities are extremely popular; yet most of them still rely primarily on text for knowledge creation and communication. This paper describes a graphical webpoll prototype—a mix of information, knowledge, and social visualization—that has been designed and deployed in an online discussion board on herbal antidepressants. The prototype, hereby named the "plot-poll", allows users to collaboratively construct a sequence of mini histograms that indicate experienced mood change during a ten week period. The pilot study shows that plot-polling is quite effective in engaging low-frequency contributors to participate. More work is required to support these findings, and reveal whether the graphic format is able to provide affective qualities to the user experience.

1. Introduction

Online communities are an extremely popular form of web activity. People use them for learning, selling, investing, gaming and a host of other purposes including just passing time. Such communities range in size from a few dozen participants, to groups that number in the thousands, and are supported by software that ranges from ephemeral chat to 3D virtual environments. In spite of their diversity of purpose, and range of forms, most communication in online communities occurs via text. In this study, we are concerned with how to encourage participation in online communities, and in particular, participation by low frequency contributors. We also hope to increase the quality of the user experience overall, not only by increasing the coherence of contributions, but through improving more difficult to measure affective qualities.

The premise of our work is that one means to this end is to provide users with non- or less-textual ways of participating. Specifically, we believe that this can be achieved through a mix of information, knowledge, and social visualization. In this paper we describe a visualization prototype—the plot-poll—that has been designed, implemented and tested in a pilot deployment in an existing online community. We begin by outlining the underpinnings of this work, briefly summarizing research on online communities, and then situating our approach to visualization relative to existing work. Next we describe the community for which the visualization was designed. After that, we describe the visualization itself, discussing the rationale behind its design and related issues. Finally, we outline its deployment, and the visualization's reception and use by the community. We conclude by reflecting on our experience, and discussing future directions.

2. Background

2.1. Online communities

Most online communities use text as the primary (and often only) means of communication. While text is powerful, we believe that providing additional modes of communication could increase the engagement and expressiveness of online environments. Efforts to engage users rarely go beyond the purely functional, such as the karma points in Slashdot.org [4] or stimulating participation through conversation channelling [13]. And, as Boyd et al. have remarked, "Nothing about the format differentiates a support newsgroup about HIV from a discussion group about the Simpsons TV show" [1, p.1]. Many researchers echo this sentiment, and have looked at the field of environmental design for inspiration. Wenger speaks of inviting different levels of participation in online communities by "building benches along the sidelines and a fire in the centre of the community that will draw people to its heat" [23]. Indeed, current interface technology can support a multiplicity of views where the audience is far more integral to the experience [14]. The premise behind this study is that engagement in online communities can be achieved through a mix of information, knowledge, and social visualization.

Wenger outlines eight categories of online communities of practice: knowledge worker's desktop;

collaborative project spaces; website communities; discussion groups; synchronous interaction tools; e-learning spaces; expertise exchanges; and knowledge bases [23]. While we hope that our approach is applicable to all eight categories, in this paper the focus is on the design of online discussion boards. To our knowledge, online discussions have not employed the sort of graphical representation we discuss in this paper.

2.2. Collaborative Knowledge Visualization

Information Visualization (IV) is the use of computer-supported, interactive, visual representations of abstract data to amplify cognition [3]. IV does this by reducing the amount of time to search and enhancing the detections of patterns, for example. A well-known IV web-based prototype is the Map of the (Stock) Market, accessible at www.smartmoney.com/marketmap.

While IV is about access and presentation of large data sets, Knowledge Visualization (KV) aims to improve the creation of knowledge among people by giving them richer graphic means of expressing what they know [2]. One example of an interactive KV is the *Knowledge Explorer*. Embedded in the user's desktop, it gives access to an Intranet document repository via a concept map each user constructs through his or her own actions [16]. This contributes to the simultaneous construction of a second, aggregate shared network map, which is also visualized. The *Knowledge Explorer* is a form of user-created knowledge, or 'social computation'.

This brings us to Social Visualizations, intuitive depictions of interactions in online space [5]. Erickson and Kellogg [7] have introduced the concept of Social Translucence with social proxies such as Babble, and more recently the TaskProxy (see Figure 1), aiming to recreate a shared sense of context in online activities [8]. Evaluations have indicated that users appreciate having a coherent, compact view of their community.

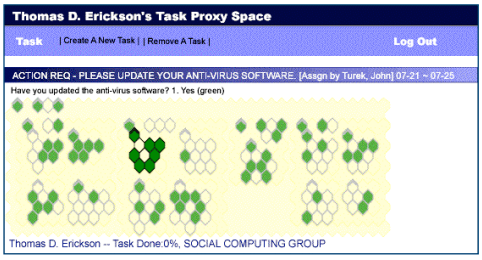

Figure 1. A social visualization

2.3. Knowledge in online discussions

People have greater motivation to perform if they are given information, via computing technology, about how their performance compares with the performance of others [9]. Electronic voting has already been explored for decades within the Delphi research. Turoff et al. [21] propose collaboratively built visual representations of concepts which users can discuss and agree upon regarding their structure and properties. Prototypes of this sort have been developed for physical discussion room settings [22], and for Intranet Group Support Systems requiring the help of a facilitator. For instance, the *Ideaquarium* [6] is a research prototype for the collaborative generation and assessment of ideas in an advertising agency. Employees enter their ideas, and rate those of others, inside an aquarium by using different graphics of fish.

The real challenge in visualization, as Douglas Engelbart has pointed out [22], is *distributed asynchronous* collaborative knowledge construction. Knowledge creation is a spiralling process of interactions between explicit and tacit knowledge [15]. Nonaka draws on the concept of "Ba"—a shared context that harbours meaning—suggesting that shared spaces (physical, virtual, or mental) can serve as a platform for knowledge creation. But current online discussion boards are hardly what Nonaka had in mind. As Scardamalia and Bereiter point out:

> Browser-dominant threaded discussion leads to a shallower, linear landscape of ideas [with no way to] signal the rising status for improved ideas as contrasted with their nondescript entity in threads, folders, and repositories where they are lost amid information glut" [19, p.6].

Web polling in online discussion boards is not as helpful to users as it could be. Typically created as a new thread, polls appear inside it, and get buried quickly. In addition, web polls typically make use of bar or pie charts, which are not suitable for visualizing multivariate data. This problem becomes evident in the site chosen for the research setting for this study.

2.4. The St. John's Wort discussion forum

The site chosen as the research setting for this study is a popular discussion forum on the herbal antidepressant St. John's Wort (SJW). The SJW herb extraction has been clinically proven to be almost as effective as traditional antidepressants [17]. The discussion board, located at http://www.sjwinfo.org, has been running for ten years and contains thousands of user comments in the form of threaded discussions. Yet, like other discussion boards, it offers no possibility for users to graphically express information. The only systematised knowledge appears as an extensive FAQ section. Visitors ask questions, such as "when will the effects of the herb start to kick in?" Hence the motivation of this study: to provide a qualitative

answer to this question using an online format that is more accessible to the user. The hundreds of contributors, and many more lurkers, at the SJW site are expected to serve as the 'population' to answer the above question. Within the past year, the number of unique visitors is a six-digit figure (evidenced at http://www.sjwinfo.org/webstats/). While there are no statistics how many of the SJW site users are currently taking the herb, it is clear from the posts that many of them are, as is also typical of depression support groups online [12].

3. Design of the plot-poll

Guidance for designing online polls is limited to suggestions that computation of the aggregated results are performed automatically, and presented to users as soon as possible in the same application for increased transparency [18]. The first step in any visualization design is to transform the raw data into a data map, and determine whether a visual structure would improve the textual information display [3]. The poll, embedded in the top of the SJW site pages represents three variables: 1) the time for the herb to start working (measured by a discrete scale of ten weeks); 2) the amount that one's mood has improved (or worsened); and 3) the number of users that have reported (marked) a particular relationship between the above variables. In this case, the time-series and mood scales are ordinal, while the frequency count is quantitative. The rule of thumb for anti-depressants, including SJW, is that one should wait at least four weeks before expecting results, but that results could be noticed as early as two weeks or as late as ten [17]. Hence, the ten week timeline.

All this information could indeed be presented more clearly with a graphical visual structure that enhances pattern recognition (see Figure 2). Tufte gives multiple examples of how plots can be extremely effective for illustrating patterns [20]. The key question is which variable gets spatial coding at the expense of others. Space, in the case of this prototype, should arguably be devoted to showing the pattern—for when and by how much the mood improvement occurs. The users can click, or 'plot', their mood for others to see in an at-a-glance manner. The prototype is meant to be a collaborative visualization, since members of the community create it. The experience of mapping, or externalizing one's subjective experience into a public space (anonymously) is expected to provide a feeling of contribution. The plot-poll also reifies a shared context and mutual story—the experience of 'hanging in' until the medication 'kicks in'.

This qualitative webpoll uses a four-point scale, emphasizing the expected positive change. The representation is meant to record the relative state of users at the current week only. The progression, however, should be evident in the aggregate, as multiple users plot their states. The plot-poll records votes from users who have been taking the herb for anywhere from one to ten weeks. To make sure we capture a representative snapshot of this period, users are requested to plot their mood only in the current week. Only one vote per person is allowed, after which the plot-poll is no longer click-able. The latest talley is visible upon refresh.

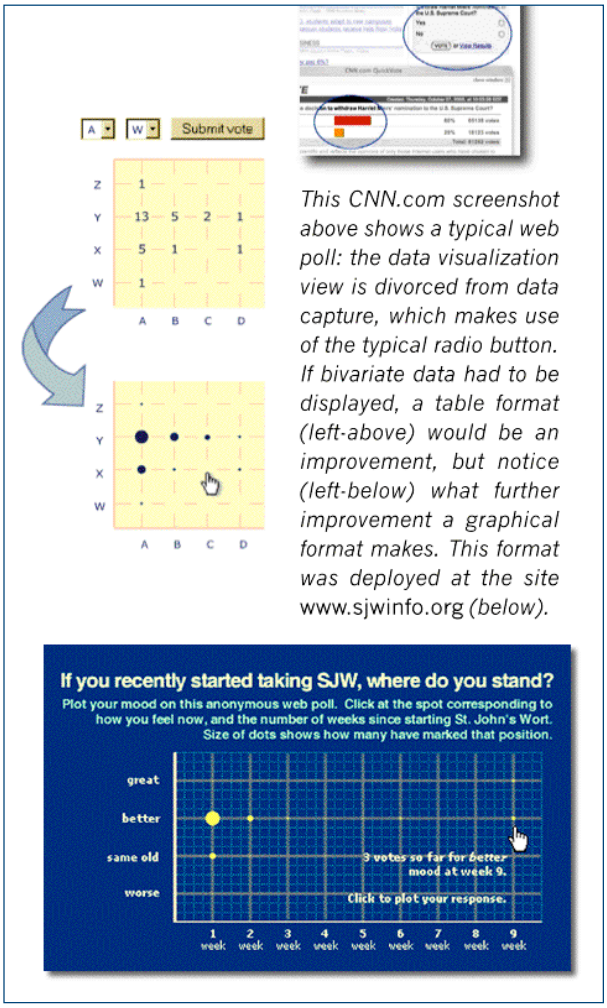

Figure 2. 2005 plot-poll version and rationale

The very first prototype of the plot-poll, pilot tested in 2005 [11], made use of dots (bubbles) that increased in size as users clicked on them (see Figure 2). Each click by a unique user increased the dot area by one pixel in diameter. A vector-based precise scaling using Flash was also considered. This type of scaling can be seen in the demo available at www.sfu.ca/~aivanov/demo.htm.

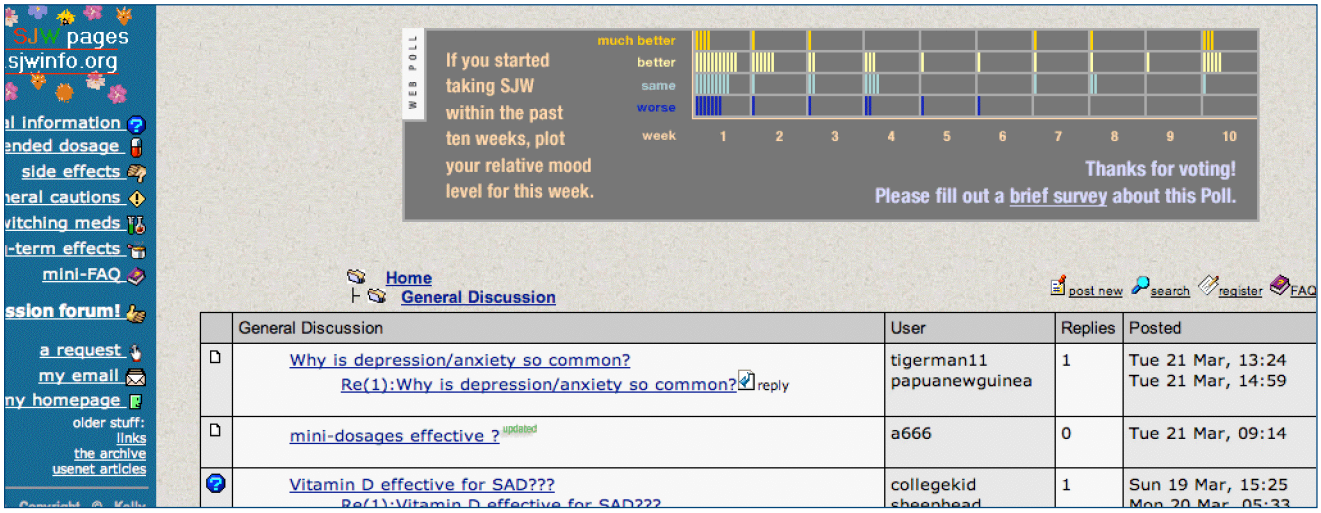

Figure 3. The 2006 plot-poll in context at day seven (screen-captured with permission)

In the current study (conducted March 2006), a different prototype was developed (see Figure 3). Our concern was that in the original approach—where each click/mark 'fused' with the previous votes (see Figure 2)—once a dozen or so users had contributed, successive clicks by other users would have an increasingly small effects on the size of bubble, thus making the act of participation at best unsatisfying, and at worst invisible. Instead, we decided to use a visual representation in which each click left a perceptible and equally-sized mark, stacking a 'brick' next to those of other users. As in the 2005 version, we believe the blending of information *capture* with *visualization* into one spot, in almost real-time, (click = mark) is a useful solution. That is, when a user clicks, he or she immediately sees the results.

Our design of the plot-poll aptly makes use of graphics for each week, which can be seen as inverted histograms. In addition, because of the narrow banner-type format, hues are used to represent the 4-point scale. Colours are blue (for cold, night or sadness, i.e. depression), leading to yellow (for warmth, brightness, i.e. happiness) for 'much better' mood.

The technology of the current plot-poll also differs from the previous version, which used the ASP.net platform. To ensure compatibility across browsers, which was a problem in the 2005 version, this time HTML is generated by a CGI script written in Perl. PNG images for the 'bricks' are programmatically generated by a second Perl script utilizing the GD library. The script stores cumulative vote counts in a file on the server. The script sets a cookie to identify users that have already submitted a vote. These users are subsequently returned a canvas that is static (not click-able), displaying a 'Thank You' note, and encourages them to fill out the survey. A prototype demo of the current version is available at http://www.collabographics.com/cgi-bin/polldemo.pl.

Other formats were also considered: line graph, Chernoff faces, and the 100% stacked bar chart. We lack space to discuss the reasons for discarding the first two, but will disucss the 100% stacked bar chart. The representation of the frequency variable is one problem, although Figure 4 illustrates a way of representing this with bar thickness. However, the format is not only more difficult to implement, but it relies entirely on colours to show a pattern—a problem for those who are colour-blind or have monitors with different calibrations.

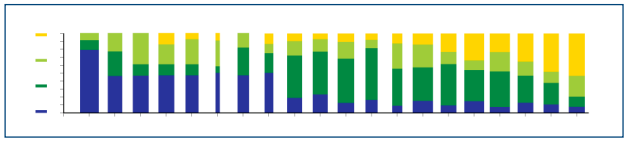

Figure 4. Mock-up of 100% stacked bar chart

4. Evaluation and discussion

Figure 3 shows the state of the plot-poll near the end of the one week test period. Seventy-six people voted in the first week of plot polling to indicate their experience with SJW. When we consider that across the three main forums that comprise the site, a total of 22 distinct users posted comments or replies to comments during the same time period, this is a very good level of participation. That is, over its first week the plot-poll has been the most popular channel for participation. The 76:22 ratio means the majority of this participation has come from lurkers, people who most likely would have been reluctant to talk, but now contributed to the site by way of the plot-poll.

In addition, an exploratory survey was administered to users to investigate their experience of the plot poll. Nine statements, on 7-point Agree/Disagree scale, were adapted from studies using the Technology Acceptance Model,

including categories for affect [10]. The statements and mean response values appear in Table 1.

Table 1. Survey items and scores; n=15

Statement	Mean
I found this poll useful.	5.27
It was easy to use.	5.80
This poll was enjoyable.	4.73
It created a sense of human contact.	4.53
Its graphics really made a difference.	4.73
I found it well designed.	4.93
I trust this poll.	5.20
I would use such a poll again.	5.93
I would use it to track my mood over time.	4.93

The survey had a 20% response rate. Respondents rated the plot-poll as useful and easy to use; most reported that they would use such a poll again. The statements regarding the affective benefits such as enjoyment were not ranked as highly, but still rate positively.

We were disappointed by two things: a lack of discussion of the plot-poll by users, and that most votes were cast for the first week. In the future, the first issue might be addressed by explicitly creating a thread for discussion of the plot-poll. The second issue deserves more elaboration. Because the plot-poll was only deployed for about a week, users were asked to respond by plotting their mood for current week (and indicating which week in the 10 week sequence their points represented). We did not allow users to enter points for multiple weeks because retrospective reporting is known to be unreliable. It is not clear whether the predominance of entries for the first week indicates that those who were most attracted to the poll were just starting out, or whether users simply responded without reading the directions, and thus started out at the 'beginning' of the chart.

Some limitations were obvious from the start, for example that web polling can not always present statistically significant information. More interesting from a visualization standpoint is portraying the number of clicks beyond what can fit within the allocated cell. In both 2005 and 2006 tests, the click limit was not reached, but one solution for a long-term deployment would be to rescale the bricks (or other markers) whenever a new order of magnitude is reached. Most importantly, future studies could compare the plot-poll with a control condition of a text-only poll. In addition, a long-term deployment that allows users to plot their mood daily and track progress over months would be very useful, as it is rated high in the survey. Aspects of mood could also be explored: such as depression versus anxiety.

5. Conclusion and next steps

Our evaluation of plot-polling is currently in the early stages. We are encouraged at the response rate during the first week of testing, and are curious to see how usage of a plot-poll format fares over time. Perhaps it will become a locus of discussion in the forum. Alternatively, poll usage in the first week may be due to its novelty, and we could see a gradual decay in usage. To further explore these issues, we intend to continue development of the design of plot-polling, for this particular use case as well as for other situations in which plot-polling might prove to be an engaging artifact for an online community. Furthermore, we plan to expand the survey items to more fully evaluate user reactions to the plot poll, but with a larger sample of users. To our knowledge, the fusion of the Technology Acceptance Model with forms of user knowledge visualization is a unique contribution.

Specifically, development of a plot-poll will be considered for online communities of Yoga practice. The mock-up shown in Figure 5 would aim to visualize knowledge on a question such as: What is the typical weight distribution in performing the Headstand pose? The plotting canvas in this case would be the triangle in the centre, representing the tripod formed by the head, arms and wrists in this pose. Users would indicate how their weight falls on the respective 'pads', which (like the bricks in the mood plot-poll) would grow to indicate 'weight'. Weight here is meant both literally (as in body weight) and figuratively (as in response significance).

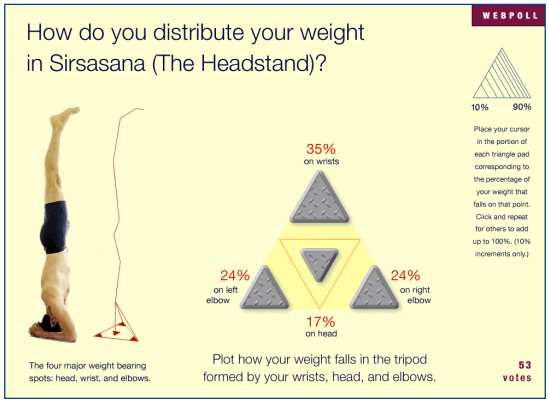

Figure 5. Mock-up of plot-poll on Yoga

To conclude, despite the preliminary nature of the evaluation, we believe that this simple and effective design of plot-polling is an example of how to increase participation in online communities. A longer-term experimental design may support this further, show which aspects of the plot-poll's design stimulate participation, and whether affective benefits are added by such collaborative knowledge visualization.

References

[1] D. Boyd, H.Y. Lee, D. Ramage, and J. Donath, "Developing Legible Visualizations for Online Social Spaces", *Proc. Hawaii Int. Conference on System Sciences*, Hawaii, January 7–10, 2002.

[2] R. Burkhard, "Towards a Framework and a Model for Knowledge Visualization", in *Knowledge Visualization and Information Visualization–Searching for Synergies,* S.O. Tergan and T. Keller (Eds.), Springer, 2005.

[3] Card, S.K., J.D. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann, Los Altos, CA, 1999.

[4] R. Cheng and J. Vassileva, "User Motivation and Persuasion Strategy for Peer-to-peer Communities", *Proc. Hawai'i Int. Conference on System Sciences,* Hawaii, 2005.

[5] J. Donath, "A Semantic Approach to Visualizing Online Conversations", *Communications of the ACM 45*(4), 2002, pp. 48-49.

[6] M. Eppler and J. Mengis, "Persistence through Visualization in Knowledge-intensive Conversations", *Proc. 9th Int. Conf. Information Visualization,* IEEE, London, UK, 2005.

[7] T. Erickson and W. Kellogg, "Social Translucence: Using Minimalist Visualizations of Social Activity to Support Collective Interaction", Readings in the Social Navigation of Information Space, K. Hook, D. Benyon, and A. Munroe (Eds.), Springer, 2002.

[8] T. Erickson, W. Huang, C. Danis, and W. Kellogg, "Social Proxy for Distributed Tasks: Design and Evaluation of a Working Prototype", *CHI'04*, Vienna, Austria, April 24–29, 2004.

[9] B.J. Fogg and S. Tseng, "Credibility and Computing Technology", *Communications of the ACM 14*(5), 1999, pp. 39-87.

[10] K. Hassanein and M. Head, "The Impact of Infusing Social Presence in the Web Interface: an Investigation across Different Products", *Int. Journal of Electronic Commerce 10*(2), Winter 2005-6, pp. 31-55.

[11] A. Ivanov and K. Funkel, "Graphical Web Polling for Online Discussion Spaces", *iLOR05 Conference* (Posters), Vancouver, Canada, November 14-17, 2005.

[12] J. Lamerichs, "Discourse of Support: Exploring Online Discussions on Depression", PhD Dissertation, Wageningen Universiteit, Netherlands, 2003.

[13] D. Millen and J. Patterson, "Stimulating Social Engagement in a Community Network", *Proc. CSCW'02*, New Orleans, USA, 2002.

[14] M. Nisenholtz, Keynote address at the 2004 Information Industry Summit Software & Information Industry Association, accessed March 20, 2006, at http://www.nytco.com/investors-presentations-20040127.html.

[15] I. Nonaka, R. Toyama, and N. Konno, "SECI, Ba and Leadership: a Unified Model of Dymanic Knowledge Creation", *Long Range Planning 33*, 2000, pp. 5-34.

[16] J. Novak and M. Wurst, "Collaborative Knowledge Visualisation for Cross-community Learning," in *Knowledge Visualization and Information Visualization–Searching for Synergies,* S.O. Tergan and T. Keller (Eds.), Springer, 2005.

[17] Pressman, A., *St. John's Wort: The Miracle Medicine*, Dell Publishing, 1998.

[18] S.R. Salz and A. Voss, "Polling in Participation Systems: an Exemplary Integration in Dito", in *DSS in the Uncertainty of the Internet Age*, T. Bui, H. Sroka, S. Stanek, and J. Gotuchowski (Eds.), University of Economics in Katowice, Slovenia, 2003, pp. 347-365.

[19] M. Scardamalia and C. Bereiter, "Knowledge Building Environments: Extending the Limits of the Possible in Education and Knowledge Work", in *Encyclopedia of Distributed Learning,* A. DiStefano, K.E. Rudestam, and R. Silverman (Eds.), Thousand Oaks, CA: Sage Publications, 2003.

[20] Tufte E., *Envisioning Information*, Cheshire, CT: Graphics Press, 1990.

[21] M. Turoff, S. Hiltz, M. Bieber, J. Fjermestad, and A. Rana, "Collaborative Discourse Structures in Computer Mediated Group Communications", *Proc. Hawaii Int. Conference on System Sciences*, Hawaii, 1999.

[22] *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making,* P.A. Kirschner, S.J. Buckingham Shum, and C.S. Carr (Eds.), Springer-Verlag: London, 2003.

[23] Wenger, E., R. McDermott, and W.M. Snyder, *Cultivating Communities of Practice*, Cambridge, MA: Harvard Business School Press, 2002.