Supporting Collaborative Deliberation Using a Large-Scale Argumentation System: The MIT Collaboratorium

Mark Klein
Center for Collective Intelligence
Massachusetts Institute of Technology
m_klein@mit.edu

Luca Iandoli
Dept. Of Business and Managerial Engineering
University of Naples Federico II
iandoli@unina.it

Current open-source/peer-production technologies, such as forums, wikis and blogs, have enabled an unprecedented explosion of global knowledge sharing, but appear to be less successful at enabling collaborative deliberation (i.e. the systematic enumeration, analysis, and selection of solution alternatives) around the complex systemic challenges, such as climate change, now facing humankind. In this paper, we present a new kind of collaboration platform, based on the large-scale application of argumentation theory, aimed at addressing this weakness. We present its rationale and design, as well as preliminary results obtained from a field test with a moderate-sized (220 member) user community.

The Challenge: Towards Internet-Enabled Collaborative Deliberation

Humankind now finds itself faced with a range of what we can call systemic problems, i.e. vastly complex challenges like climate change that affect every one of us and are affected by every one of our actions. Such problems call for us to be able to engage in effective deliberations on a global scale. The spectacular emergence of the Internet has enabled unprecedented opportunities for such interactions - via such open-source/peer-production (OSPP) tools as email, instant messaging, news groups, chat rooms, blogs, wikis, podcasts, video and picture sharing sites, and the like - on a scale that was impossible a few short years ago. To date, however, such large-scale interactions have been incoherent and dispersed, contributions vary widely in quality, and there has been no clear way to converge on well-supported decisions concerning what actions humanity should take to solve their most pressing problems. Can we do a better job of harnessing the vast collective intelligence now potentially available to us? This paper explores this question, reviewing the limitations of current technologies, and presenting a large-scale argumentation system aimed at addressing these challenges.

Limitations of Current Technologies

Let us define "collaborative deliberation", for this context, as the synergistic channeling of the efforts of many minds towards enumerating, evaluating, and coming to consensus over responses to some complex challenge (Walton and Krabbe 1995). How well does current OSPP technology achieve this goal? While such tools have been remarkably successful at enabling a global explosion of idea and knowledge sharing, they face serious shortcomings with respect to supporting collaborative deliberation (Sunstein 2006). The content captured by such tools is notorious for having a poor signal-to-noise ratio, with many repetitive and low-quality posts, especially when addressing controversial issues. Coverage of a topic is generally unsystematic, since it is created bottom-up. Group interactions are all too easily hijacked by a narrow set of "hot" issues or loud voices, leading to such phenomena as forum "flame wars" and wiki "edit wars". OSPP systems do not inherently encourage or enforce any standards concerning what constitutes valid argumentation, so postings are often bias- rather than evidence- or logic-based. Users of such systems

also tend to self-assemble into groups that share the same opinions ("balkanization"), so they see only a subset of the issues, ideas, and arguments potentially relevant to a problem.

Argumentation tools (Kirschner, Shum et al. 2005) can, we believe, help address these weaknesses in OSPP systems. They work by helping groups define networks of *issues* (questions to be answered), *ideas* (alternative answers for a question), and *arguments* (statements that support or detract an idea or other argument) where every distinct issue, idea, or argument appears just once (see Figure 1 below).

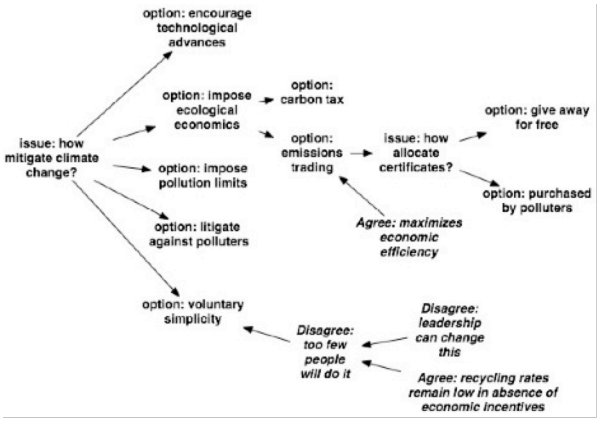

Figure 1. An example of an argument map.

Such tools help make deliberations, even complex ones, more systematic and complete. It's easy to see what has been covered to date, and what has not. All perspectives on an issue, regardless of the community the author comes from, are co-located in the argument map, working against any tendency to balkanization. Since a point can appear just once, users are not able to sidetrack the process by sheer repetition. The central role of argument entities, finally, encourages users to express the evidence and logic in favor of the ideas they favor.

Argumentation systems require, however, more skill and care to use than OSPP systems, and have as a result been used almost exclusively for small-scale physically co-located team meetings, where the participants engage in a free-form discussion while a single facilitator captures these deliberations in the form of a commonly-viewable argument map (Shum, Selvin et al. 2006). They have also been used, to a much lesser extent, to enable non-facilitated deliberations, over the Internet, with physically distributed participants (Jonassen and Jr 2005) (Chklovski, Ratnakar et al. 2005) (Lowrance, Harrison et al. 2001) (Karacapilidis, Loukis et al. 2004) (Heng and de Moor 2003). The scale of use in these cases has also been small, however, with on the order of 10 participants or so working together on any given task, far less than what is needed to address complex problems like climate change.

Towards Large-Scale Argumentation

In this paper, we suggest that we can transcend the current limitations of deliberation support technologies by creating *large-scale argumentation systems*, i.e. systems that combine the generative power of OSPP systems with the compact and systematic knowledge organization enabled by argumentation systems. Our hypothesis is that this approach, by providing a logic- rather than a time-based representation for capturing user contributions, and by encouraging evidence-based reasoning and critical thinking, will preserve the large-scale participation that characterizes OSPP systems, while qualitatively reducing the prevalence of the dysfunctions (such as low signal to noise ratios, poor argumentation, and balkanization) they currently face, thereby enabling highly effective large-scale deliberations.

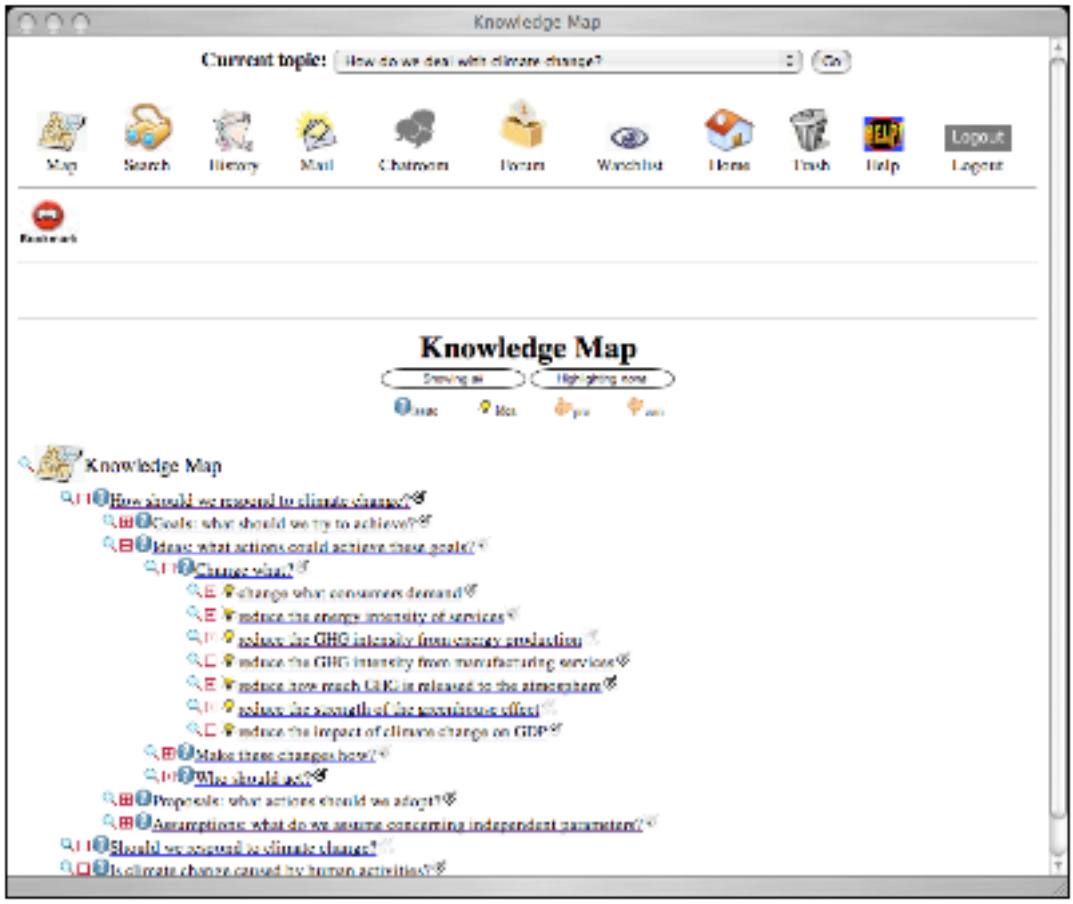

Figure 2. The main Collaboratorium display.

We have implemented a prototype web-based system, called the Collaboratorium, in order to test this hypothesis. The basic design of the system is simple. Users are asked to create a network of posts organized as an argument map (figure 2, above).

Users are expected to follow a set of simple guidelines to ensure the map is well-structured. Each post should represent a single issue, idea, or argument, and should not replicate a point that has already been made elsewhere in the argument map. It should be attached to the appropriate part of the argument map. Changing a post in order to undermine someone else's point of view is forbidden: if one disagrees with an idea or argument, the user can capture this by creating *new* posts that present their alternative ideas or counter-arguments. Because good argument-mapping skills are not universal, a class of users known as "moderators" is charged with ensuring that new posts are correctly structured. Posts are initially given a status of "pending", and can only be certified by moderators. Until a post is certified, it appears in a small font, can not be rated, nor can any posts be added to it. Moderators are expected to leave comments with the posts explaining what needs to be done in order for them to be eligible for certification. They are also expected to re-organize the map as needed to cluster related topics to ensure a moderate branching factor and thus make it easier to find content on a given subject. The Collaboratorium supports widely-used OSPP functions such as rating (so the community can identify the most important issues, promising ideas, and compelling arguments), version histories (so users can see the history of a post as well as roll it back to a previous version) and watchlists (where users are automatically notified, by email, when changes have been made to posts they are interested in).

Creating an argument map lays out the space of possible solutions to a given problem. In order to come to conclusions concerning what actions they will take with respect to that problem, community members are expected to create a distinct branch in the argument map in which the ideas each represent proposals - made up of combinations of possible solutions - that can then be rated by the community. The highest-rated proposal represents the community decision.

The key uncertainty with our approach concerns whether argument mapping will work with large scale communities. Previous argumentation systems have, as we have noted, off-loaded argument mapping to a single dedicated facilitator. But this approach clearly does not scale to thousands of users; we need to distribute the mapping activity across the user population itself. Will users take on this task? There are good reasons to believe that they will. One reason concerns communication styles. In small groups, relationship management is usually primary, so participants often prefer to communicate implicitly and indirectly, especially at first, in order to avoid unnecessary conflicts. In this context, a technology like argument mapping that forces people to make their arguments explicit can actually be a liability, rather than an asset. But this effect flips on its head at large scales. Implicit communication becomes unnecessary (and often confusing) when most people do not know each other personally, or can float potentially contentious proposals under aliases. So we can expect that resistance to argument mapping will decrease as the user community grows in size.

Other reasons concern the nature of incentives in OSPP systems. It has been found (Hars and Ou 2002) (Lakhani and Wolf 2005) (Roberts, Hann et al. 2006) (Bonaccorsi and Rossi 2004) that users contribute to such systems predominantly for two reasons: (1) to find their tribe (i.e. get connected with people who share their interests) and (2) to become a hero (have a substantive positive impact on the community they identify with). There is widespread disaffection with the low signal-to-noise ratio of current OSPP tools. It seems clear that the number of distinct issues, ideas, and arguments in a discussion will grow, after a certain point, much more slowly than the number of participants. The larger the user community, therefore, the more potential redundancy there will be, and thus the more value argument mapping offers in terms of improving the signal to noise ratio. We can thus expect that, as the system grows, users will increasingly recognize and respond to the opportunity to "become a hero" by contributing something (i.e. creating a value-rich argument map) that is highly valued by the community. Argument mapping also increases user's chances of "finding their tribe". While contributing to unstructured discussions is easier, the high volume and redundancy of such discussions means that most posts will probably be overlooked by most readers. In an argument map, by contrast, if you have a unique point to make, it has a much greater chance of being read.

Another question concerns whether the user community will be able to provide sufficient *numbers* of users with good argument mapping skills. There are several reasons to believe this will be possible. There is already a substantial world-wide community of people with argument mapping skills. One organization alone (cognexus.org) has already trained and certified hundreds of people in the IBIS methodology that underlies the Collaboratorium. Other similar argument mapping tools, such as debatemapper.com, have their own user communities that could also contribute. Argument mapping is, in addition, a natural skill for lawyers, philosophers, mathematicians, library scientists, and others who create proofs or ontologies for a living. Such individuals can probably learn an IBIS-like formalism very quickly, and may be inspired by the opportunity to contribute their critical thinking skills to debates around critical challenges like climate change, even if they do not have content expertise in that area. Another point is that, as the system scales, we can assume that the argument-mapping burden *per user* will decrease. This is because the number of mapping-savvy users should scale linearly with the overall user population, but the number of novel ideas that need to be mapped should scale (as we mentioned above) *less* than linearly. If the

Collaboratorium works like most OSPP systems, user contributions will follow a power law, so we can expect that a relatively small corps of "power users" will take on the bulk of the argument mapping tasks. Most people will just have to read or comment on, as opposed to create, argument maps.

See (Klein 2007) for further details on the design of the Collaboratorium.

Evaluation

An evaluation of the Collaboratorium was performed in December of 2007 at the University of Naples with a community of 220 graduate students, which was asked to use the Collaboratorium to deliberate, over a period of three weeks, on the topic "the future of biofuels in Italy". The students were all given, at the start, a lecture introduction on how to use the Collaboratorium. While data analysis is still ongoing, we are able to report some preliminary results.

We observed a very high level of user participation. Remarkably, the Collaboratorium was active almost continuously, except for a daily hiatus between roughly 3 and 6 am:

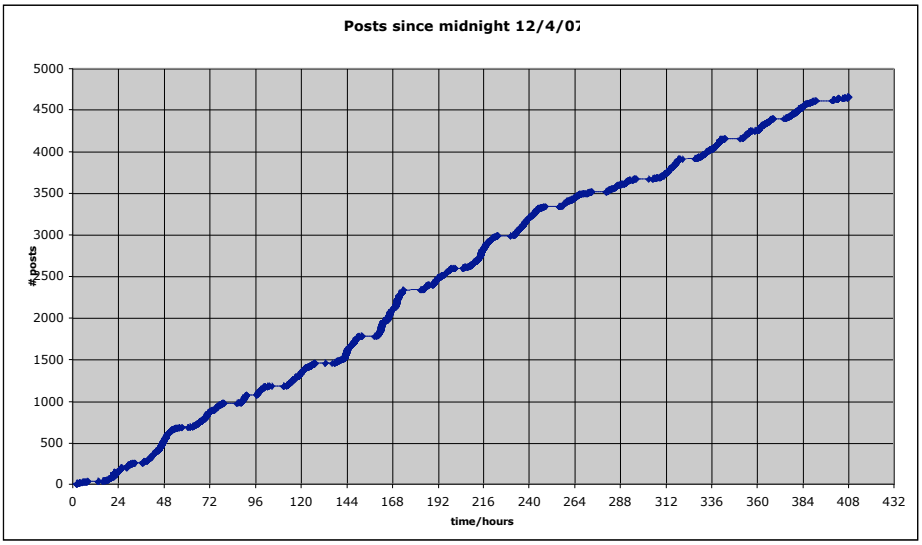

Figure 3. Growth in number of posts over time.

About 180 out of 220 users contributed at least a few posts, and the most active contributed 40 or more posts each. In two weeks the students posted nearly 3000 issues ideas and arguments (of which roughly 1900 were eventually certified) in addition to over 2000 comments (table 1):

Type of Post	**Number of Posts**	**Number of Certified Posts**	**% (certified)**
Issue	242	89	5%
Idea	962	452	24%
Pro	1488	1045	55%
Con	402	325	17%
Comment	2009	n/a	n/a
Grand total	5003	1911	100%

There were, however, relatively few ratings: each post received an average of only 2.2 ratings, often including one from the post author.

The intensity of participation varied widely among users (Figure 4), roughly following the power law distribution that has been found to be typical of most on-line communities.

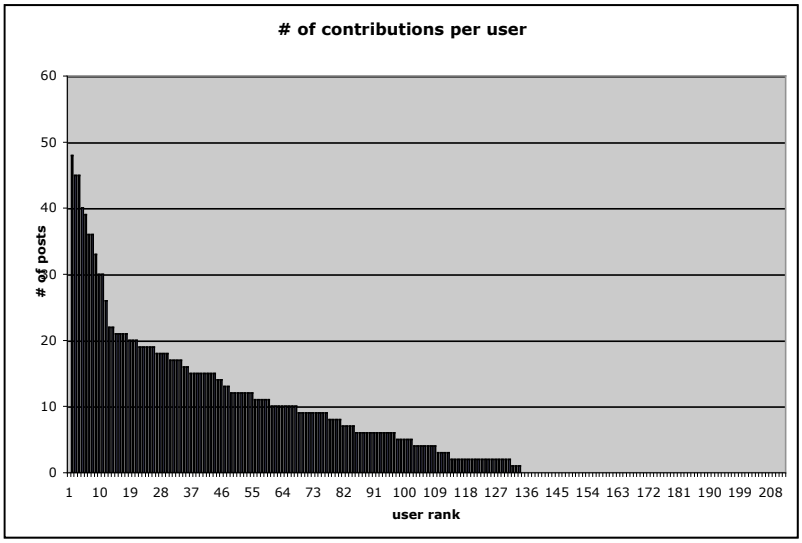

Figure 4. Distribution of number of posts per user

The breadth and depth of coverage was, in our judgment, quite good: this non-expert community of students was able to create a remarkably comprehensive map of the current debate on biofuels, complete with many references to statistics and publications, exploring everything from technology and policy issues to environmental, economic and socio-political impacts. Moreover, the proportion of out-of-topic posts was negligible – about 0.1%.

Though students' participation may have been influenced by their perception that the experiment was a course task for which they could be evaluated by their professor, their informal face-to-face and on-line comments, posted on the Collaboratorium as well as on a threaded discussion forum run independently by a students association web site, showed that they found the experiment interesting and appreciated the innovative characteristics of the Collaboratorium. However, to assess user satisfaction more carefully, we have designed a survey, which is currently in progress.

As expected, at the beginning of the experiment many users did not fully grasp the argumentation formalism. Many users adopted, rather, a kind of forum frame in which they tended to publish news-like posts (e.g. "France creates incentives for biofuels") rather than distinct issues, ideas or arguments. Other common mistakes included: not distinguishing between ideas and arguments, putting multiple arguments into a single argument post, linking arguments to a logically irrelevant location in the argument map, and proliferating questions and ideas without any associated pro/con arguments. After a while we observed an improvement in the use of the platform, as users developed confidence, profited from moderator feedback, and learned to use the tool.

The degree of debate was significant. 70% of all arguments were attached to posts authored by someone else. The great majority of all arguments (again, 70%) were pros rather than cons, however, and the *depth* of the argument trees was relatively small:

Depth of argument tree	% of all arguments
1	85%
2	12%
3	2%
4	1%

Table 2. Depth of argument trees in Naples evaluation.

This relative dearth of extended argument chains may have been an outcome of the students' reluctance to criticize the contributions of their peers, and thus may be an artifact of the co-located nature of the user population. Other possible explanations include: i) inertia deriving from the predominant use of forums and wikis, ii) lack of adequate incentives for debate, which induced users to overemphasize authorship of novel ideas, iii) the short time window compared to the learning curve of users with the new tools, iv) the lack of specific expertise and motivation of the students on the topic at hand, leading to fast content saturation and inability to explore specific subtopics in full depth.

Important lessons were learned concerning community governance. Moderators played a crucial role. They supported users with comments and suggestions and, by ensuring a logically-organized argument map, helped users rapidly locate the contexts where their piece of knowledge can best be placed. For these reasons it is crucial to have enough moderators working to ensure fast certification and timely reorganization of the argument map. With the existing data we can roughly estimate the requisite number of moderators per users. A cadre of from 2 to 5 moderators (the number varied from day to day according to their other commitments) was able to more or less keep up with 180 active authors, but only by dint of an unsustainably heavy investment of their time. We estimate that a more realistic time commitment would require that roughly 5% of the active users also act as moderators.

Many more lessons almost certainly remain to be gleaned from the test dataset. The Collaboratorium software recorded essentially every user interaction with the knowledge base, including every view or modification or rating of any post, so we have a complete time-stamped record of the evolution of the argument map and what the users did while creating it, a database of over 110,000 distinct events. We also conducted pre- and post-evaluation tests of the bio-fuels content expertise and critical reasoning skills of all the participants. A thorough analysis of this data will be presented in future publications.

Conclusions and Next Steps

In this paper we have presented a new collaborative deliberation platform called the Collaboratorium. The key contribution of the Collaboratorium is that it combines the logic-, rather than time-based knowledge organization structure of argumentation systems with the immense generative power of large-scale, open-source/peer production systems. This combination, we have argued, offers the promise of enabling qualitatively more productive large-scale collaborative deliberations. Our initial results appear to support this. The Naples evaluation resulted in the creation of what is to our knowledge one of the largest argument maps ever built, on a complex topic, over the course of two weeks, working with over 200 novice users.

Our next steps will include a side-by-side comparison between the Collaboratorium and other OSPP (wiki, forum) technologies, in order to assess their relative strengths and weaknesses. The Naples evaluation involved a relatively small number of users, by Internet standards, and the incentives structure was distorted, in all likelihood, by the fact that they were students in a class as well as co-located peers.

The experiment also ran, perforce, over a limited time window. Further evaluations will aim to remove these artificial constraints by assessing the platform with much larger, and truly open, user communities. Increased scale will probably require changes in design choices and user incentives. Among the most critical improvements we underline: designing mechanisms and rules able to generate a self-organized hierarchy of user roles (authors, and moderators), improving the information visualization & retrieval elements of the platform, providing on-line support to users (such as on line help and training tools), and building tools to increase moderator productivity. We are currently identifying other possible contexts for assessing and applying the Collaboratorium, ranging from problem solving within organizations to pedagogy with student populations.

Acknowledgements

I would like to gratefully acknowledge the many valuable contributions to the ideas described herein made by Marco Cioffi, Kara Penn and David Price.

References

Bonaccorsi, A. and C. Rossi (2004) "Altruistic individuals, selfish firms? The structure of motivation in Open Source software." DOI:

Chklovski, T., V. Ratnakar and Y. Gil (2005). User interfaces with semi-formal representations: a study of designing argumentation structures. Proceedings of the 10th international conference on Intelligent user interfaces: 130-136.

Hars, A. and S. Ou (2002). Working for Free? Motivations for Participating in Open-Source Projects. International Journal of Electronic Commerce 6(3): 25 - 39.

Heng, M. S. H. and A. de Moor (2003). From Habermas's communicative theory to practice on the internet. Information Systems Journal 13(4): 331-352.

Jonassen, D. and H. R. Jr (2005). Mapping alternative discourse structures onto computer conferences. International Journal of Knowledge and Learning 1(1/2): 113-129.

Karacapilidis, N., E. Loukis and S. Dimopoulos (2004). A Web-Based System for Supporting Structured Collaboration in the Public Sector. Lecture Notes in Computer Science: 218-225.

Kirschner, P. A., S. J. B. Shum and C. S. C. Eds (2005). Visualizing Argumentation: Software tools for collaborative and educational sense-making. Information Visualization 4: 59-60.

Klein, M. (2007). The MIT Collaboratorium: Enabling Effective Large-Scale Deliberation for Complex Problems. Research Paper 4679-08. MIT Sloan School of Management. http://ssrn.com/abstract=1085295

Lakhani, K. R. and R. G. Wolf (2005). Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. Perspectives on Free and Open Source Software. J. Feller, B. Fitzgerald, S. Hissam and K. R. Lakhani, MIT Press.

Lowrance, J. D., I. W. Harrison and A. C. Rodriguez (2001). Capturing Analytic Thought. First International Conference on Knowledge Capture: 84-91.

Roberts, J., I.-H. Hann and S. Slaughter (2006). Understanding the Motivations, Participation and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects. Management science 52(7): 984-999.

Shum, S. J. B., A. M. Selvin, M. Sierhuis, J. Conklin and C. B. Haley (2006). Hypermedia Support for Argumentation-Based Rationale: 15 Years on from gIBIS and QOC. Rationale Management in Software Engineering. A. H. Dutoit, R. McCall, I. Mistrik and B. Paech, Springer-Verlag.

Sunstein, C. R. (2006). Infotopia: How Many Minds Produce Knowledge, Oxford University Press.

Walton, D. N. and E. C. W. Krabbe (1995). Commitment in dialogue: Basic concepts of interpersonal reasoning. Albany, NY, State University of New York Press.