

A First Step toward Scaling-up Deliberation: Optimizing Large Group E-Deliberation using Argument Maps

Paolo Spada¹, Mark Klein², Raffaele Calabretta³, Luca Iandoli⁴, Ivana Quinto⁵, and all the members of the online community of Insieme per il PD.

Abstract⁶

The most successful democratic innovations have been able to promote good deliberation by employing small discussion groups (5 to 15 participants). In this paper we show that it is possible to design online democratic innovations that can induce good argumentation in large groups of more than 150 participants. The essay is based on a field experiment implemented in a real life democratic innovation designed to promote good e-deliberation and organized in collaboration with the largest online community of the Italian Democratic Party (Insieme per il PD). The field experiment aims to explore the possibility of improving the quality of e-deliberation by comparing the performance of a traditional forum with the performance of an argument mapping software structured to optimize argumentation (Deliberatorium). With respect to a traditional forum technology the argument mapping software 1) does not significantly reduce participation, 2) reduces the total number of ideas generated by the users, and 3) increases the exchange of arguments in favor of and against each idea. Given the centrality of good argumentation in promoting the quality of deliberation, this experiment shows a promising first step in scaling-up deliberation.

¹ Centre for the Study of Democratic Institution, University of British Columbia.

² MIT Center for Collective Intelligence.

³ Institute of Cognitive Sciences and Technologies of the Italian National Research Council.

⁴ University of Naples Federico II.

⁵ University of Naples Federico II.

⁶ We want to thank Donald Green and Holger Kern for critical contributions to the experimental design, Archon Fung, John Gastil, Quinton Mayne, Hollie Russon-Gilman, and all the members of the Ash Center Democracy Seminar for important feedback on the early drafts of this essay. Many thanks also to on. Sandro Gozi, Giuseppe Rotondo, Marco Lombardo and Gabriele Maestri and all the other members of Insieme per il PD that allowed us complete access to their online community, without their help this research would not have been possible. Finally many thanks go to Anna De Liddo, Ivana Quinto, Ali Gurkan, Ralf Groetket, Alessandro Rossi and Vincenzo Lanzetta for the assistance in moderating the discussion.

1. INTRODUCTION

Is it possible to scale-up good deliberation online? Or can good deliberation be achieved only in face to face small group discussions (5 to 15 people)? Over the course of the last thirty years researchers and practitioners have managed to refine the technology of face to face small group deliberation using a combination of theoretical insights, trial and error, and ad hoc measures of success. Deliberative Polls, Consensus Conferences, America Speaks' 21st Century Town Hall Meetings, Citizens' Juries, Planning Cells, National Citizens' Assemblies⁷ are just a few examples of democratic innovations that emerged via this alchemic process and are considered best practices of face to face deliberation (Fung 2006; Warren and Pearse 2008; Smith 2009; Mansbridge 2010). These innovations employ similar sequences of devices (e.g., quasi random sampling, small group moderated discussions, balanced provision of information, Borda voting mechanisms, etc.) whose individual effect is still not completely clear, but that combined have shown to promote some of the aspects of what we consider good deliberation (Black et al. 2010; De Vries et al. 2011). Relying on many integrated small group face to face discussions implies that such structured processes are complex and expensive, both for the organizers and the participants, limiting their frequency of adoption and scalability to communities of a few hundred (Chadwick 2008).

⁷ For a repository of case studies and methodologies of Democratic Innovations see the Participedia Project: <http://participedia.net/>

With the advent of the internet many hoped that asynchronous text based discussions⁸ could generate new venues of high quality deliberation that could easily scale up. But, as the literature on the quality of online discussions has well documented, these new spaces rarely fulfill such promises (Sunstein 2006; Kies 2010; Hartz-Karp and Sullivan 2014).

However the current comparison between online text-based discussions and structured face to face deliberations that are the results of years of experimentation on how to promote good deliberation is problematic. Online text-based deliberations, depending on the specific venue, are more similar to the spontaneous discussion in a thematic club, in a work group that has to develop a product, in a large assembly, or in a pub around the news of the day (Coleman 2004). Recent developments in crowdsourcing and idea-sharing software have harnessed the potential of large scale collaborations, but have not yet solved the problem of online text based discussions (Klein 2011). All this software was not designed with the specific purpose of promoting good deliberation, thus it is not surprising that it does not achieve such outcome.

What would happen if an online venue of deliberation was designed specifically to promote the exchange of rational arguments? Is it possible to overcome the well-known problems of text-based online deliberation? Is it possible to achieve good deliberation in large groups using online tools?

In this paper we explore these questions by presenting the results of a pilot field experiment within an online democratic innovation designed specifically to promote good deliberation. The

⁸ This is the type of discussion that occurs on twitter, forums, chat lines, wiki editorial discussions, and in general in any software that offers the possibility of leaving comments. The more general topic of computer mediated interactions is beyond the scope of this paper.

experiment did not involve a large set of participants (640), but employed large deliberation groups each composed of 160 citizens, more than ten times the size of the average small groups (5 to 15 participants) used in face to face deliberations. If we can achieve good e-deliberation in groups containing 160 participants then the possibility of scaling-up deliberation to an actual crowd of thousands of people appears less remote.

The online democratic innovation was implemented in the spring of 2012 in collaboration with a group within the Italian Democratic Party, *Insieme per il PD* (Together for the Democratic Party).⁹ This democratic innovation translates online a participatory and deliberative decision making process (Intra-party Deliberative Referendums – “Doparie”) designed to deepen parties’ internal democracy (Calabretta 2011).

The topic of discussion was the electoral law, which was, and still is, highly controversial and continually discussed by the Italian media due to the dysfunctional reform introduced by Silvio Berlusconi in 2005, and due to the rise of a new political force (Five Star Movement) that upset the bipolar environment that has characterized the Italian polity during the nineties (De Rosa 2013; Turner 2013; Lanzone and Rombi 2014). The Deliberative Referendum embedded a pilot experimental analysis to investigate the effect of different structures of large group e-deliberation.¹⁰ Two randomly generated groups of 160 participants were assigned to discuss using the Deliberatorium, an argument mapping software (Klein, Cioffi, and Malone 2007;

⁹ <http://insiemeperilpd.ning.com/>

¹⁰ For recent experiments varying the structure of online forums see Rhee and Kim 2009; Smith, John, and Sturgis 2013.

Schneider, Groza, and Passant 2013), while two randomly generated groups of 160 participants were assigned to discuss using a standard forum.

Our objective in this essay is three fold. First, we want to evaluate whether restricting the structure of discussion using an argument mapping software affects retention rate and the activity level of the users. Second, we want to compare the quality of deliberation generated by the traditional forum and by the argument map. Third, we want to compare the organizational cost of the two competing mechanisms. If the argument map is written mostly by the moderators it cannot be a good technology to scale up deliberation.

Foreshadowing our main results, our study uncovered an important trade-off. The more structured the discussion becomes the more argumentation occurs, but the less ideas are generated. Thus we do not recommend the exclusive use of argument mapping software for small group discussions whose objective is brainstorming. The employment of argument mapping software is recommended for large scale online consultations in which the proliferation of ideas is easy to achieve, while argumentation is really difficult to sustain.

Our experiment shows that argument mapping not only induces better analysis of each idea (+140% arguments per idea) without affecting participants' retention rate, but it is also significantly cheaper to manage (42 vs 163 man-hours). Most importantly for practitioners the Deliberatorium eliminates the necessity of time consuming and sometimes non-transparent experts post-processing. This experiment sets an important precedent for scaling up deliberation.

The rest of the paper is organized as follows: in section two we briefly overview the literature on the limits of existing online deliberation technologies, in section three we introduce large scale argumentation as a e-deliberation method, in section four we describe the experimental

design and the various metrics we employ. Section five presents the results of the experiment and section six concludes by offering insights for academics and practitioners.

2. STRENGTHS AND LIMITATIONS OF EXISTING ONLINE DELIBERATION TECHNOLOGIES

One important aspect of achieving good deliberation is the careful weighing of the merits of a broad range of solutions (Habermas 1984; Cohen 1989; Elster 1998). There are many other important facets of the quality of deliberation (Bächtiger, Niemeyer, et al. 2010), but *ceteris paribus*, if we can increase the quality of argumentation in a group discussion the overall quality of deliberation will most likely improve.

How well do existing online technologies meet the challenge of promoting good argumentation? A wide range of social computing technologies have emerged in the past few decades, including Wikis (e.g., Wikipedia), opensource software development efforts (e.g., Linux), solution competitions (e.g, Innocentive.com), design contest (e.g, Threadless), idea-sharing systems (e.g, ideastorm.com), peer-filtering sites (e.g, Slashdot), group decision support systems (e.g, meetingsphere.com), review websites (e.g., yelp.com), and Q&A websites (e.g., Stack Overflow). Experience with such systems has shown that they can foster voluntary contributions and participation at a vast scale, which in turn can lead to remarkably powerful emergent collaborative phenomena (Surowiecki 2005; Sunstein 2006; Tapscott and Williams 2008; Page 2008; Doan, Ramakrishnan, and Halevy 2011; Coleman and Shane 2012).

But while we have progressed in the field of e-collaborations, the technology that is used to conduct discussions online has not really changed in the past thirty years and is based on asynchronous text based interactions. These large group e-discussion are prone to dysfunctional

emergent behaviors¹¹ that can deeply undercut the quality of the deliberation outcomes. Problems such as stronger group polarization (Sia, Tan, and Wei 2002; Sunstein 2006; Chen 2013), informational cascades (Hansen, Hendricks, and Rendsvig 2013), low signal-to-noise ratio (Cotton and Yorke 2006), dysfunctional argumentation, information overload (Losee 1989), scattered content, redundancy, non-collaborativeness (Klein 2011) plague asynchronous text based interactions.

Due to the widespread use of these technologies the problems of existing forms of online discussion are well known even beyond academia and have generated folk laws such as the infamous Godwin Rule of Nazi Analogies.¹² Thus the prevailing sentiment in the practitioners' community is that there is something special about face to face structured deliberation. Body language and body chemistry generate a set of signals that can be translated online only partially. This structural difference combined with the low quality of deliberation observed in existing online venues of discussion fuels the belief that face to face deliberation is intrinsically better than online deliberation. In turn this belief greatly influences the type of venues of deliberations that are adopted. Most existing democratic innovations reflect the idea that the quality of online deliberation is inevitably low and thus implement mainly vertical mechanisms of interaction between the organizers and the participants (e.g., e-survey, i-voting¹³, provision of information,

¹¹ An emergent behavior or emergent property can appear when a number of simple entities (agents) operate in an environment, forming more complex behaviors as a collective.

¹² https://en.wikipedia.org/wiki/Godwin%27s_law

¹³ Internet voting, i.e. the subset of e-voting technologies that specializes on remote voting systems via internet.

geo-mapping) without promoting significant horizontal interactions among participants (Schlosberg, D., Shulman, S. W., and Zvestoski, S. 2006; Allegretti 2011). This concatenation of events generates a vicious cycle that reinforces the existing beliefs of the low quality of online deliberations (Spada and Allegretti 2014).

The approach we introduce in the next section was specifically designed to overcome these problems by having the participants cogenerate an argument map together with moderators (Klein, Cioffi, and Malone 2007; Klein and Iandoli 2008; Klein 2011).

3. LARGE SCALE ARGUMENTATION

Large-scale argumentation represents a promising approach to addressing the weaknesses with current deliberation technologies. Argumentation tools (Kirschner, Buckingham-Shum, and Carr 2003; Moor and Aakhus 2006; Walton 2006; Klein, Cioffi, and Malone 2007; Schneider, Groza, and Passant 2013) take an argument-centric approach based on allowing groups to systematically capture their deliberations as tree structures made up of issues (questions to be answered), ideas (possible answers for a question), and arguments (statements that support or detract from an idea or argument) that define a space of possible solutions to a given problem.

Figure 1: An example of an argument map approximately here

Such tools have many advantages (Kirschner, Buckingham-Shum, and Carr 2003). Every unique point appears just once, radically increasing the signal-to-noise ratio, and all content on a given question is co-located, making it easy to find what has or has not been said on any topic, fostering more systematic and complete coverage, and counteracting balkanization by putting all competing ideas and arguments right next to each other (Suthers et al. 2008; Gürkan et al. 2010).

Careful critical thinking is promoted, because users are implicitly encouraged to express the evidence and logic in favor of the options they prefer (van Gelder 2007; Okada 2008). Users, finally, can collaboratively refine proposed solutions. One user can, for example, propose an idea, a second raise an issue concerning how some aspect of that idea can be implemented, and a third propose possible resolutions for that issue.

Most argumentation systems have been used by individuals or in small-scale settings, relying in the latter case on a facilitator to capture the free-form interactions of a collocated group as a commonly-viewable argument map (Shum et al. 2006). Argumentation systems have also been used, to a much lesser extent, to enable distributed deliberations over the Internet. These maps tend to be poorly structured, however, because many users are not skilled argument mappers, and the scale of participation has been small, typically involving only a handful of authors on any given task (Schneider, Groza, and Passant 2013).

The argument mapping software employed in this experiment (Deliberatorium) overcomes some of the pre-existing limitations by introducing moderators tasked to review and certify each post. It uses the IBIS argumentation formalism (Conklin 2006), which is one of the simplest and most widely applied. The moderators' job is part education, and part quality control. Posts, when initially created, are given a "pending" status and can only be viewed by other moderators. If a post doesn't adequately follow the argument map conventions, moderators will either fix it or leave comments explaining what needs to be done. Once a moderator has verified that a post follows the conventions, the post is "certified" and becomes available to be viewed, edited, commented on, or rated by the general user population. The certification process helps ensure well-structured maps, and provides incentives for users to learn the argument formalism. Moderators serve a relatively modest role in all this: their role is not to evaluate the merits of a

post, but simply to work with authors to ensure that the content is structured in a way that maximizes its utility to the community at large.

Before this experiment the Deliberatorium had only been tested with students in class experiments (Gürkan et al. 2010). Even with the assistance of moderators the Deliberatorium remains a fairly complex system to use in comparison to a standard forum.

In order to better explore its feasibility in non-academic environments we designed, together with a group within the Italian Democratic Party, an online deliberative decision-making system based on the Intra-party Deliberative Referendums model. Then we used such environment to investigate experimentally the advantages and disadvantages of the Deliberatorium with respect to the more traditional forum technology. This research design sacrifices the perfect control of the environment that can be achieved in the lab, for a more realistic evaluation of what would occur in real life when a party implements a Deliberative Referendum.¹⁴

4. USING THE LESSONS OF FACE TO FACE DELIBERATION BEST PRACTICES TO IMPROVE ONLINE DELIBERATION

The Intra-party Deliberative Referendums (“Doparie”) are democratic innovations aimed at improving political parties’ internal decision making (Calabretta 2011). They are consultations of party members/supporters to be held on key and divisive issues during the period between one

¹⁴ For an introduction on the advantages and disadvantages of field experiments see Gerber and Green 2012.

election and the next.¹⁵ Three key elements distinguish the Deliberative Referendum from traditional discussion within parties.

1. The Intra-party Deliberative Referendum is a multi-event democratic innovation characterized first by a deliberative phase and then by a decision making phase that uses a voting mechanism.
2. The Intra-party Deliberative Referendum requires provision of balanced information to all participants.
3. The Intra-party Deliberative Referendum adopts procedures to promote good deliberation.

Before this experiment the Intra-party Deliberative Referendum was just a proof of concept discussed within the parties and civil society organization of the Italian left (Calabretta 2011). Therefore, for this experiment we designed an Intra-party Deliberative Referendum using devices frequently adopted by existing best practices.

Similarly to Citizens' Assemblies we created a time-limited process of enrollment that collected information on the participants (Warren and Pearse 2008). Differently from Citizens' Assemblies we did not discard any participant to create a small representative sample of the population. Using the enrollment survey we created four randomly generated discussion groups of 160 participants each. Following recent results in the lab experimental literature on small group deliberation showing the important effects of gender group composition on participation (Karpowitz, Mendelberg, and Shaker 2012) we balanced the four discussion groups across gender, age, education, intensity of preferences over the two main families of electoral laws

¹⁵ The term "Doparie" is a neologism, a compound term formed from the Italian words "dopo" ('after') and "primarie" ('primaries') (Calabretta 2011).

(majoritarian vs proportional), interest in the topic, membership in the party, vote in the past primaries for the party leader, internet usage, and an additional question that aimed to distinguish the degree of skepticism on online discussions. Following the examples of successful implementations of participatory budgeting — one of the most widely adopted participatory democratic innovations — we shaped an advertising campaign aimed at highlighting the democracy enhancing characteristics of the process. The campaign, conducted via social media, blogs and the websites of *Insieme per il PD* and sympathizers, centered on a balanced message that was not in favor of one or the other potential solutions, but that highlighted the possibility of bringing a larger diversity of ideas to the attention of the hierarchy of the party.¹⁶ Given that one of the common problems of participatory budgeting is the mismanagement of expectations (Allegretti 2011), the campaign paid particular attention to sending a clear message about the limited potential impact of the process over the politics of the entire party. *Insieme per il PD* is a relatively small group within the democratic party.

Following the lessons of recent multi-channel engagement processes the campaign highlighted the multiplicity of ways a user could participate, ranking them in increasing order of effort. Following the example of the briefing package included in deliberative polls we introduced a capacity building briefing video that explained the process and a virtual library containing balanced information on a large number of electoral systems (Isernia and Fishkin 2014). Participants themselves were invited to contribute to the virtual library by sending material via

¹⁶ The promo video (<https://www.youtube.com/watch?v=noMaHClrrbU>) for example stressed the possibility of changing the Democratic Party to make it more representative of the base.

email as long as the information was purely descriptive and did not contain any persuasive message.¹⁷

The process was divided in three stages, advertising and enrollment phase (1 month), deliberation (3 weeks) and voting (1 week). Participants were asked to contribute to all three. Many posted information about the process in their own blogs and Facebook pages. In the final i-voting phase participants were allowed to change their vote within the allotted time, but could cast only one final vote. Security of the vote was monitored by allowing only one vote for enrolled user. In the enrollment phase the ip-address and the email were automatically checked and a subsequent manual screening was conducted. A team of nine moderators monitored the discussions and provided help with the platforms. An email reminder system contacted participants when one of their contributions to the discussion had received a rating or a comment. Users could rate each post, but each user could see only his own personal rating to prevent bandwagon effects.

The objective of these features was to translate online some of the key devices of successful face to face democratic innovations (e.g., balance information package, moderators, balanced groups, balanced experts) and avoid some of their most common pitfalls (e.g., mismanagement of expectations, low legitimacy due to a non-neutral stance of the organizers that clearly support one of the solutions, high cost of participation). Jointly these features created a civil space of discussion. In one month our moderators did not have to ban a single participant for disruptive behavior or inappropriate language.

¹⁷ An Italian mathematician sent his custom electoral system through this method.

5. CAN WE DO BETTER? AN EXPERIMENTAL DESIGN TO MEASURE THE QUALITIES OF DIFFERENT STRUCTURES OF E-DELIBERATION

The democratic innovation we designed constructed a respectful environment conducive for deliberation, but could not solve by itself the problems generated by online asynchronous text based discussion. In order to explore how to optimize the Intra-party Deliberative Referendum for online discussions, we assigned two groups of 160 participants to discuss using a forum tool and two groups of 160 participants to discuss using the Deliberatorium. Each discussion lasted three weeks. Using a pre-survey, groups were randomly generated and balanced across a variety of individual characteristics that might have affected the propensity to participate online. The two tools were designed to appear as similar as possible to reduce the chance that any performance differences we find are due to spurious differences in the user interfaces. For this reason instead of using a widely available forum format (e.g., phpBB) we created our own forum from a stripped down version of the Deliberatorium. The interfaces were visually very similar and provided, for example, the same set of basic functions, including a chat room to discuss off topic and ask for help from moderators, a suggestions tool that lists posts that may be of interest to the users, a search tool for finding posts with given keywords, an activity tool that gives information on how active the discussion has been, a profile tool that allows users to change their passwords, a virtual library with background information on electoral reform, a suggestions box where they could leave feedback on how to improve the system. There was also a help facility that included text and brief videos on how to use the system, along with sections on topics such as privacy and use of the experimental data.

Forum users could add comments to forums, create new sub-forums, and rate other people's posts. Map users could add new issues, ideas, or arguments to the map, edit posts they created earlier, leave suggestions on other authors' posts, and rate posts. In the forum comments were

immediately visible, in the map user contributions were invisible until they were certified by a moderator.

Map moderators were responsible for checking and certifying posts, as described above, while forum moderators just made sure that the posts were free of inappropriate content such as spam or abusive language. Users, in all conditions, received email, every two or three days, notifying them of recent activity in their discussion areas and encouraging them to view and contribute to them. Both the web forums and argument maps listed three identical topics for discussion: “What electoral law should Italy adopt?”, “What other important questions should be asked about the electoral system?” and “What topics should be deliberated about in the future?” Users could make whatever contributions they wanted within these fixed macro-issues.

The evaluation ran for three weeks, from April 1, 2012 to April 21, 2012. After that, users were given a brief post-survey that asked about their experiences with the systems, as well as about their opinions on the major electoral reform alternatives enumerated during the deliberation. About three-quarters of the active participants filled out the post survey. In addition to these surveys, the system comprehensively tracked all user actions in a time-stamped event log. In total, close to 36,000 user events were logged.

Given the difference between an argument map and a forum using traditional direct methods of comparison, such as civility of discussion, integrative complexity (Brundidge et al. 2014), the Deliberation Quality Index (Steenbergen et al. 2003; Bächtiger, Shikano, et al. 2010)¹⁸, or

¹⁸ The original DQI has four categories of coding: free participation (i.e. presence of interruptions), level of justification, content of the justification, respect, and constructive politics. In an asynchronous text based online interaction the first does not apply. In argument map

measures of noise, would automatically generate better results for the map or the forum by construction.

Each post in the map has to be certified by a moderator and has to be justified. Noise, e.g., off topic posts, is not certified and thus is not visible to users.¹⁹ The map does not offer the space for

respect is automatically enforced by the moderators. Similarly, in an argument map there is no way to detect significant variation in the level of constructive politics. Due to the specific type of deliberation targeted to members and supporters of the Italian Democratic party and on the topic of an electoral law reform all the arguments proposed were generally justified in terms of providing a better electoral law for the country or the party and thus they would not generate enough variance in content of justification component of the index. We are left with the level of justification that is a modified measure of integrative complexity. Thus overall the original DQI is not particularly helpful to compare our two treatments.

¹⁹ In forums 1 & 2 there was a total of 2,717 and 1,322 words that were categorized as “non-content posts” during the post processing phase. The non-content category comprises meta posts about the organization of the discussion, posts that are difficult to understand, posts that are unrelated to the discussion, and duplicate posts. All these posts were eliminated in the post-processing phase. The self-generated map had no noise, but it had a few posts that were not certified by the end of the process, more precisely in map 1 there were 11 ideas, 8 pros, 3 cons that were not certified, while in map 2 there were 14 ideas, 2 pros, 7 cons, and 6 issues that were not certified. Some of these elements were tests from the users, some other elements were actual posts that required some clarification from the users.

non-constructive or disrespectful speech acts. Thus indexes that measure noise and civility of the discussion, everything else equal, would automatically rank the map higher.

The map isolates concepts and connects them visually, eliminating the necessity of complex speech acts. Thus a forum post such as: *“I think that a pure proportional system does not work for Italy because it generates too many parties, I believe that we should introduce, instead, a majoritarian system similar to France because it limits the proliferation of small parties, generates stable coalitions and promotes more accountability.”* in map form would be split in two separate ideas (pure proportional system; France majoritarian system), one argument against the first idea (it generates too many parties), three arguments in favor the second idea (it limits the proliferation of parties, it generates stable coalitions, promotes more accountability).

Figure 2: Map representation approximately here

Thus measures of integrative complexity based on word counts of integrating signifiers (e.g., “the reason is”) applied to each element of the map separately would detect low complexity because the map employs graphical connections to integrate its various elements. Thus the forum would score automatically higher in integrative complexity.

In the map the moderator dialogues with the users helping them structure the post according to the IBIS formalism (Conklin 2006). Note that if a participant provides an argument in favor/against an idea that contain a story, an example, or an expression of preference (I love idea X, because it is the best), such post is not eliminated, but is moved in the comment section that pertains to the appropriate element of the map (Idea X). If the appropriate element in the map does not exist, the user is invited to create it by the moderator.

The moderator also can intervene directly to fix posts if the user is not online. A typical intervention consists in splitting the post up and then pointing to the community in the chatline what the moderator has done so that other users can learn how to improve their posts, thus reducing the intervention of moderators. The final map does not contain any noise or redundancy, just a set of non-certified elements.

Given these issues with existing approaches in this essay, we compare the topology of the argument map generated by the users and the moderators in the Deliberatorium treatment with the best possible map that a team of experts of argument mapping could construct on the basis of the raw data of the forum treatment.²⁰ In order to reduce differences between the two treatments the moderators of the map and the forum were the same, while the authors of this paper post-processed the forums. Post-processing the forums' results not only has the advantage of generating another argument map that can be directly compared with the Deliberatorium's map, but it is also similar to the process that occurs in real life after an online consultation process. Raw ideas provided by the community of users via a forum are filtered and processed by a team of community engagement experts.

To compare the qualities of the maps generated by the two treatments we investigate measures such as the number of ideas generated, the number of arguments, the number of arguments per idea, the depth of the map and the depth of the ideas and sub-ideas, and the depth of the arguments. These measure are meant to explore which tool is more capable of gathering

²⁰ Note that the experts are experts of argument mapping and not experts of the discussed topic. These experts are not allowed to add content to the discussion, but only to reorganize it in a map. The coding rules are available upon request.

ideas and exploring issues and arguments in favor of or against an idea, i.e. our measure focus on the argumentation aspect of deliberative performance (Mercier and Landemore 2012).

6. OUTCOMES

Of the 640 enrolled participants, 373 (just over 50%) logged into the website at least once. The cumulative number of first-time users increased over time and we had new users entering the software right to the last day. We interrupted the process at 21 days due to the research design,²¹ but if the deliberation had continued participation probably would have risen further (Figures 3a and 3b). The increase in participation over time might have something to do with positive feedback effects generated by the quality of deliberation, but we do not have the data nor the proper research design to investigate such hypothesis.

Figure 3a: Cumulative number of first login per day

Figure 3b: Cumulative number of first login per day

Figures 3a and 3b show the importance of allowing enough time for proper deliberation online. Our experiment lasted three weeks and does not show a plateau in the number of first-time logins. Further studies are required to understand the optimal length of this processes and when participation plateaus on the basis of different groups' size, intensity of advertising and type of deliberation. We had minimal resources for advertising, thus these graphs likely represent a lower bound of what can be achieved in similar processes within parties.

²¹ The experimental design was approved by Harvard University IRB process.

These graphs also highlight one of the difficulties of using this type of e-deliberation to construct indirect measures of deliberation quality based on changes of participants' preferences and attitudes (Black et al. 2010). The treatment effects of asynchronous text-based deliberation on participants level variable (e.g., change in preferences) are heterogeneous and depend on the time of the first login and the time a participant spends on the platform. Existing experiments investigating individual level effects control for overall noncompliance, but do not adjust for heterogeneous treatment effects of this sort (Smith, John, and Sturgis 2013).

The two treatments generated very similar average levels of activity at the group level. Figure 4 shows that the average number of participant per day per treatment was slightly greater than 10 (row 6), and that among the enrolled 160 participants slightly more than 50% concluded the process participating in the vote (row 5). Thus the overall retention and activity level is not affected by the structure of discussion.

When we investigate the other major event type (Edit a post, Create a post, Rate a post, View a post, View the virtual library) we observe differences only in the rate a post category. The argument map increases the number of ratings by 37.5, i.e. a 220% increase in ratings with respect the forum average of 17 ($p=0.001$). Note that this measure does not depend on our coding. We hypothesize that rating was more frequent in maps than forums because, in the forum, they are less meaningful; a forum post may make several points, so a rating thereof is potentially ambiguous, while a map post always makes only a single point. Note that the ratings we employed were not publicly visible, thus this effect is not influenced by social pressure. This is a first signal that shows how the re-organization of content in an argument map might elicit significant behavioral effects.

Figure 4: number of unique users who performed each major event type at least once.

When we look at the overall activity in the four platforms we see that 194 participants, or approximately 30% of the 640 enrolled, created one or more posts. If we consider the level of activity among those that tried the software at least once (373), we see that more than half of them created content (52%). Given the 1/9/90 rule of thumb of internet participation these results are actually surprising and exceed our expectations.²² This unusual level of participation might have been generated by a combination of factors including the controversial topic that was concurrently being covered by the media,²³ the design of the Doparia that targets the supporters of a political party that have a shared value system, the self-selected sample, and the design of the process itself, a novelty effect.

²² The rule of thumb states that in a collaborative website such as a wiki, 90% of the participants of a community only view content, 9% of the participants edit content, and 1% of the participants actively create new content.

²³ During 2011 a large nonpartisan popular movement collected more than a million notarized signatures to request an abrogative referendum on the electoral law. If the referendum would have passed, Italy would have gone back to the law introduced in 1993 eliminating all the changes introduced by Silvio Berlusconi in 2005. On January 24th, 2012 the Constitutional Court of Italy established that this referendum was not admissible because it would have generated a “normative vacuum.” Many interpreted this move as politically motivated to protect the newly installed technical government of Mario Monti that was charged to create a new electoral law. The media attention around these controversial decisions was high.

When we look at the number of unique users in each condition over the course of the 21 days we again find minimal difference across the four groups, with maximum peaks of around 25 users, and minimum at around 5 users (results available upon request).

We conclude from this pilot experiment that the employment of argument mapping software is a viable substitute for online forums that has no significant impact on retention or activity levels. This is a surprising result given that the map appears to require more effort from the users. We hypothesize that the increased cost of participation might be compensated by the gain in organizational structure and clarity that the maps generate over time.

However, this unusual level of sustained participation might also be an effect of the novelty of the tool and the fact that the users knew they were participating in research. Only a widespread adoption of argument mapping software can actually separate these effects.

In order to compare the argumentation quality of the forum with the one of the map we recoded each forum in an argument map form. The post-processing required more than 160 man-hours. The moderation of the forum was minimal because the discussion was extremely civil and required less than three man-hours which mostly involved reading posts. The moderation of the map was more intense in the first week and then slowly required less work due to the fact that users had learned how to use the platform and the map facilitated by the help chatline which offered numerous examples of common errors and mistakes. In total we spent 42 man-hours for the map conditions. Overall the traditional e-consultation method, forum plus expert post-processing, required almost four times more man-hours than the maps (163 vs 42).

Comparing the self-generated map with the forum post-processed by a team of argument mapping experts (Figure 5), we observe that the forums created more than three time as many ideas as the user-generated map (290 vs 78). The difference of the average effect of the forums

and the maps on the number of ideas generated (106) is statistically significant ($p=0.013$). Thus the map decreases the number of ideas by 73%. Note that while this is clearly a problem in small group discussions, in large group discussions in which thousands of ideas that are minor variations of each other are posted, it might be an advantage.

The number of pros, cons and issues generated by the forum is larger than those generated by the map, but the difference of the average effects is not significant. Furthermore if we factor in that most of the pros in the forum conditions are generated by the same author that provided the idea in the first place (“preaching”) then we see that the maps generate almost as many “other authors” pros than the forum condition (50 vs 51).

Figure 5: comparing the two treatments approximately here

In order to capture in relative terms the effect of the treatments, we investigate the following argumentation indexes:

$$\textit{Argumentation index 1} = \frac{(\textit{pros} + \textit{cons} + \textit{issues})}{\textit{ideas}}$$

$$\textit{Argumentation index 2} = \frac{(\textit{pros} + \textit{cons} + \textit{issues} - \textit{preaching})}{\textit{ideas}}$$

The first index captures the amount of argumentation per idea independently of who has generated the arguments, while the second index, analyzes more specifically the amount of exchange of arguments per idea among different users by eliminating pros to an idea that were generated by the same author of the idea (preaching). The difference in the average effect of the map and the forum on index 1 is 136 percentage points, i.e., the self-generated map creates more than twice as many arguments per idea than the forum post-processed by the experts; this result

is statistically significant ($p=0.041$). Thus the map increases the first index by 140%. The difference in the average effect of the map and the forum on index 2 is slightly larger, 138 percentage points, this result is again statistically significant ($p=0.036$). Thus the map increases the second index by 214%. These two indexes highlight a key difference between the two treatments: the map promotes a richer exchange of arguments per idea than the forum.

Not surprisingly when we look at the average depth²⁴ and average breadth²⁵ (Figure 6) we find a significant difference between the maps generated by the experts and the ones generated by the users. The map generated by the experts shows a significant increase in depth of 32 percentage points ($p=0.01$) and a significant increase in the breadth of 104 percentage points ($p=0.02$). These results are driven by the multiplication of ideas and sub-ideas that increase the layers of the map generated on the basis of the forum.

Figure 6: Topology approximately here

²⁴ The depth of an element of the map represents the layer of the map in which the element is located. The top layer issue assume value zero, each layer below increases the depth by one. For example if a map is composed of an issue, an idea to solve such issue, an argument in favor of such idea, and an issue of such argument, the depth of the last issue would be three.

²⁵ The breadth of an element of the map represents the number of direct children that such element has in the layer immediately below. For example if a map is composed of an issue, and two ideas to solve such issue, and an argument in favor of one of those ideas, the breadth of the issue would be two (the two ideas).

However when we look at the depth of the arguments (A_{depth}),²⁶ excluding ideas and sub-ideas, we find that the user-generated map shows a non significant increase in the depth of the arguments of 19 percentage points ($p=0.11$). Again this is a signal that the map generated by the users can potentially create more indepth discussion of each idea. The effect however is not large enough to be statistically significant in our small sample of four groups. Further replications of this experiment with a larger set of groups are required to better identify this effect.

7. DISCUSSION

This field experiment shows the concrete possibility of inducing good deliberation in large groups of 160 participants. Existing applications of e-deliberation are based on forum and chat technologies and show significant problems (e.g., scattered content, low signal-to-noise ratio, balkanization and dysfunctional argumentation). The result of this experiment shows that a carefully structured e-deliberation employing argument mapping software allows us to overcome such problems; content is organized, noise is filtered, and balkanization is impossible by construction, and while participants propose fewer ideas, they do evaluate them more rigorously.

It is important to note that the Deliberatorium is just a technology, and it does not guarantee to achieve the results that we induced in our experiment by itself. The careful design of the Intra-party Deliberative Referendum (Doparia) contributed critically to the quality of the e-deliberation. Our experiment is not designed to separate the effect of the Deliberatorium from the effect of the Intra-party Deliberative Referendum (Doparia). What our experiment shows is that we can use the Deliberatorium to improve a carefully designed e-deliberation process by

²⁶ The depth of the arguments of the map (A_{depth}) captures the depth of the chain of pros and cons assuming that the first pro or con has depth zero independent of its actual depth.

inducing better argumentation than traditional threaded forum technology. Compared to the forum, the Deliberatorium increases the number of arguments per idea by 136 percentage points.

To our great surprise, in this experiment the employment of a restricted structure of discussion does not affect users' retention rate nor the users' daily average activity. More research is needed to uncover the precise mechanism that generated this counterintuitive result.

The argument map significantly reduces, however, the quantity of ideas posted by the users, thus it should be used with care in small group discussions.

Given that the most common problems of large online discussions are the proliferation of many similar ideas and the lack of argumentation, the Deliberatorium appears to be a good instrument to solve both problems at the same time *without reducing participation*.

Another important advantage of the argument map, with respect to traditional online discussion technologies, is that it does not rely on experts to generate immediately usable information. Most e-consultations collect large quantities of raw data that require months to be analyzed by experts before they can be synthesized in actionable proposals. This filtering process is non-transparent and often generates frustration in the users that see their ideas changed or disregarded without sufficient explanation. The argument map combines the idea-generation phase with the redundancy/noise elimination phase. Thus, users receive immediate and transparent feedback with regard to ideas that cannot be included in the map because they are redundant or off topic. In this incarnation the moderators' policy was pre-set, and non-certified posts were not visible to the users, but nothing prohibits introducing a meta-discussion about such rules to identify the best governance policy before the actual e-deliberation begins. This meta-discussions are successfully implemented by many participatory budgeting programs around the world.

While theoretically the optimal group size for argument maps is unlimited, further research is required to uncover potential problems with crowdsourcing argument maps. It is possible that information overload occurs again when the number of participants is greater than a certain threshold. But even in the worst case scenario in which the optimal size is the one we adopted (160 participants), that would imply, at least, a 10-fold expansion of the size of face to face small-groups adopted in current deliberations.

It is also important to note that the results we are presenting are conditional to the sample of self-selected participants. The democratic innovation that we designed was open to the public and our advertising targeted all supporters of the Italian Democratic Party and all citizens interested in pushing the Democratic Party to adopt Deliberative Referendums. Most existing crowd-sourcing processes do not involve a representative sample of the population, and that is why our field experiment was not designed to create a representative sample. But the technology to create representative samples of the population exists and is routinely used by large surveys and mini-publics. What had been difficult, before this experiment, was to promote good deliberation in large groups independently of their online or offline mode of interaction.

This research constitutes a first step in constructing an online environment that is capable of sustaining good deliberation involving thousands of people.

FIGURES

Figure 1: An example of an argument map



Figure 2: Map representation



Figure 3a: Cumulative number of first login per day

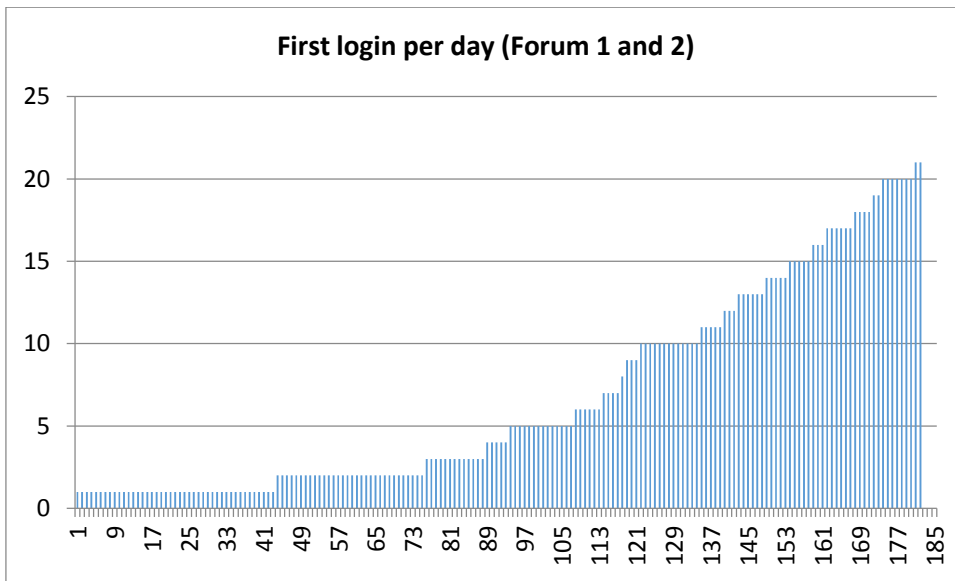


Figure 3b: Cumulative number of first login per day

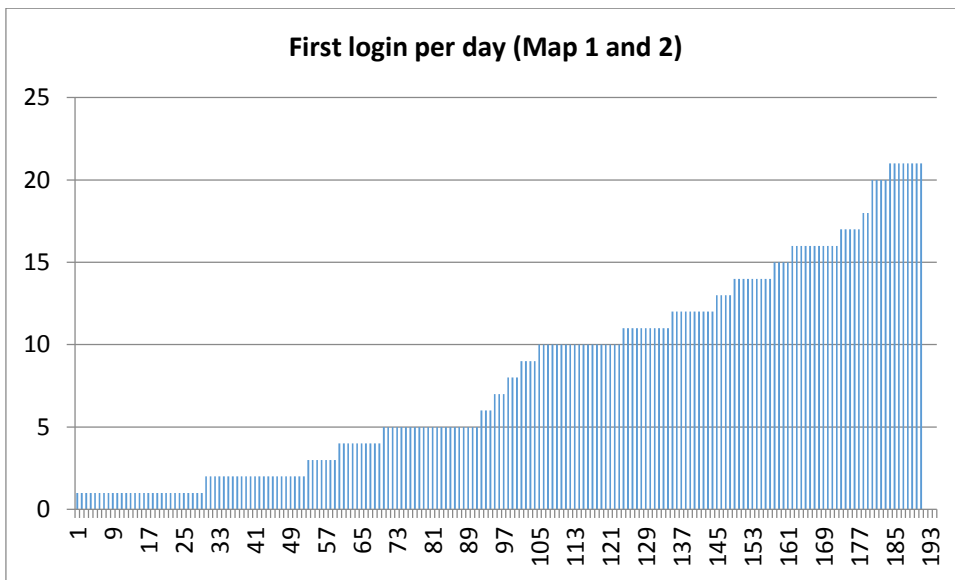


Figure 4: number of unique users who performed each major event type at least once.

Event Type	Forum 1	Forum 2	Map 1	Map 2
EDIT	40	47	50	42
CREATE	41	47	57	49
RATE	16	18	54	55
VIEW	79	88	87	87
VOTE (Final Ballot)	83	89	83	84
VIEW RESOURCES (virtual library)	25	30	29	31
Average number of unique users per day	11.2	13	13.5	13.8

Figure 5: comparing the two treatments²⁷

Group	Ideas	Pros	Preaching	Cons	Issues	Arg. Index 1	Arg. Index 2
Forum 1 & experts	155	91	50	84	12	1.21	0.89
Forum 2 & experts	135	53	43	43	1	0.72	0.4
Map 1	33	33	9	32	7	2.18	1.91
Map 2	45	41	15	59	11	2.47	2.13

²⁷ We consider only elements that have been cogenerated with the users and that have been certified. This excludes ideas and issues in the forum conditions that were generated by experts to organize the map and have no actual direct reference in the forum, and excludes elements of the self-generated map that were not certified by the end of the process.

Figure 6: topology

Group	Avg. depth	Std. depth	Avg. breadth	Std. breadth	Avg. Adepth	Std. Adepth
Forum 1 & experts	3.56	0.98	3.81	4.48	0.19	0.49
Forum 2 & experts	3.51	0.93	4.01	4.42	0.06	0.24
Map 1	3.19	0.96	2.98	2.96	0.34	0.59
Map 2	3.23	0.83	2.77	3.25	0.29	0.57

REFERENCES

- Allegretti, Giovanni. 2011. "From Skepticism to Mutual Support: Towards a Structural Change in the Relations between Participatory Budgeting and the Information and Communication Technologies?" *Working Paper IVR World Congress*.
- Bächtiger, André, Simon Niemeyer, Michael Neblo, Marco R. Steenbergen, and Jürg Steiner. 2010. "Disentangling Diversity in Deliberative Democracy: Competing Theories, Their Blind Spots and Complementarities*." *Journal of Political Philosophy* 18 (1): 32–63. doi:10.1111/j.1467-9760.2009.00342.x.
- Bächtiger, André, Susumu Shikano, Seraina Pedrini, and Mirjam Ryser Ryser. 2010. *Measuring Deliberation 2.0: Standards, Discourse Types, and Sequenzialization*. <http://ash.harvard.edu/extension/ash/docs/baechtiger.pdf>.
- Black, L. W., S. Burkhalter, J. Gastil, and J. Stromer-Galley. 2010. "Methods for Analyzing and Measuring Group Deliberation." In *The Sourcebook for Political Communication Research: Methods, Measures, and Analytic Techniques*, edited by E. P. Bucy and R. L. Holbert, 323–45. New York, NY: Routledge.
- Brundidge, Jennifer, Scott A. Reid, Sujin Choi, and Ashley Muddiman. 2014. "The 'Deliberative Digital Divide': Opinion Leadership and Integrative Complexity in the U.S. Political Blogosphere: Opinion Leadership in Political Blogs." *Political Psychology*, June, n/a – n/a. doi:10.1111/pops.12201.
- Calabretta, R. 2011. "Doparies: A Novel Party Deliberative and Aggregative Decision-Making Mechanism to Improve the Quality of Representative Democracy." *SAGE Open* 1 (3). doi:10.1177/2158244011433699.
- Chadwick, Andrew. 2008. "Web 2.0: New Challenges for the Study of E-Democracy in an Era of Informational Exuberance." *ISJLP* 5: 9.
- Chen, Hsuanwei Michelle. 2013. "Group Polarization in Virtual Communities: The Case of Stock Message Boards." <https://www.ideals.illinois.edu/handle/2142/36051>.
- Cohen, Joshua. 1989. "Deliberation and Democratic Legitimacy." *Debates in Contemporary Political Philosophy*, 342.
- Coleman, Stephen. 2004. "Connecting Parliament to the Public via the Internet: Two Case Studies of Online Consultations." *Information, Communication & Society* 7 (1): 1–22.
- Coleman, Stephen, and Peter M. Shane. 2012. *Connecting Democracy: Online Consultation and the Flow of Political Communication*. MIT Press.
- Conklin, Jeffrey. 2006. *Dialogue Mapping: Building Shared Understanding of Wicked Problems*. Wiley.
- Cotton, Deborah, and Jon Yorke. 2006. "Analysing Online Discussions: What Are Students Learning." In *23rd Annual Ascilite Conference: Who's Learning*. http://www.ascilite.org/conferences/sydney06/proceeding/pdf_papers/p65.pdf.
- De Rosa, Rosanna. 2013. "Voice of the People or Cybercratic Centralism?" In *Conference for E-Democracy and Open Governement*, 89. <http://www.diva-portal.org/smash/get/diva2:660884/FULLTEXT01.pdf#page=90>.
- De Vries, Raymond, Aimee E. Stanczyk, Kerry A. Ryan, and Scott Y. H. Kim. 2011. "A Framework for Assessing the Quality of Democratic Deliberation: Enhancing Deliberation as a Tool For Bioethics." *Journal of Empirical Research on Human Research Ethics: An International Journal* 6 (3): 3–17. doi:10.1525/jer.2011.6.3.3.
- Doan, Anhui, Raghu Ramakrishnan, and Alon Y. Halevy. 2011. "Crowdsourcing Systems on the World-Wide Web." *Communications of the ACM* 54 (4): 86. doi:10.1145/1924421.1924442.
- Elster, Jon. 1998. *Deliberative Democracy*. Cambridge University Press.
- Fung, Archon. 2006. "Varieties of Participation in Complex Governance." *Public Administration Review* 66 (s1): 66–75.

- Ganuza, Ernesto, Gianpaolo Baiocchi, and others. 2012. "The Power of Ambiguity: How Participatory Budgeting Travels the Globe." <http://digital.csic.es/handle/10261/66210>.
- Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. W. W. Norton.
- Gürkan, Ali, Luca Iandoli, Mark Klein, and Giuseppe Zollo. 2010. "Mediating Debate through on-Line Large-Scale Argumentation: Evidence from the Field." *Information Sciences* 180 (19): 3686–3702. doi:10.1016/j.ins.2010.06.011.
- Habermas, Jürgen. 1984. *The Theory of Communicative Action: Reason and the Rationalization of Society*. Beacon Press.
- Hansen, Pelle G., Vincent F. Hendricks, and Rasmus K. Rendsvig. 2013. "Infostorms." *Metaphilosophy* 44 (3): 301–26.
- Hartz-Karp, Janette, and Brian Sullivan. 2014. "The Unfulfilled Promise of Online Deliberation." *Journal of Public Deliberation* 10 (1): 16.
- Isernia, P., and J. S. Fishkin. 2014. "The EuroPolis Deliberative Poll." *European Union Politics* 15 (3): 311–27. doi:10.1177/1465116514531508.
- Karpowitz, Christopher F., Tali Mendelberg, and Lee Shaker. 2012. "Gender Inequality in Deliberative Participation." *American Political Science Review* 106 (03): 533–47. doi:10.1017/S0003055412000329.
- Kies, Raphaël. 2010. *Promises and Limits of Web-Deliberation*. Palgrave Macmillan.
- Kirschner, Paul A., Simon J. Buckingham-Shum, and Chad S. Carr. 2003. *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*. Springer Science & Business Media.
- Klein, Mark. 2011. "How to Harvest Collective Wisdom on Complex Problems: An Introduction to the MIT Deliberatorium." *Center for Collective Intelligence Working Paper*. http://cci.mit.edu/docs/working_papers_2012_2013/kleinwp2013.pdf.
- Klein, Mark, M. Cioffi, and T. Malone. 2007. "Achieving Collective Intelligence via Large-Scale On-Line Argumentation." In *ICIW*, 58. <http://cci.mit.edu/publications/CCLwp2007-01.pdf>.
- Klein, Mark, and Luca Iandoli. 2008. "Supporting Collaborative Deliberation Using a Large-Scale Argumentation System: The MIT Collaboratorium Paper 239."
- Lanzone, Maria Elisabetta, and Stefano Rombi. 2014. "Who Did Participate in the Online Primary Elections of the Five Star Movement (M5S) in Italy? Causes, Features and Effects of the Selection Process." *PARTECIPAZIONE E CONFLITTO* 7 (1): 170–91.
- Losee, R. M. 1989. "Minimizing Information Overload: The Ranking of Electronic Messages." *Journal of Information Science* 15 (3): 179–89. doi:10.1177/016555158901500306.
- Mansbridge, Jane. 2010. "Deliberative Polling as the Gold Standard." *The Good Society* 19 (1): 55–62. doi:10.1353/gso.0.0085.
- Mercier, Hugo, and Hélène Landemore. 2012. "Reasoning Is for Arguing: Understanding the Successes and Failures of Deliberation: Reasoning Is for Arguing." *Political Psychology* 33 (2): 243–58. doi:10.1111/j.1467-9221.2012.00873.x.
- Moor, Aldo de, and Mark Aakhus. 2006. "Argumentation Support: From Technologies to Tools." *Communications of the ACM* 49 (3): 93–98.
- Okada, Alexandra. 2008. "Scaffolding School Pupils' Scientific Argumentation with Evidence-Based Dialogue Maps." In *Knowledge Cartography*, 131–62. Springer. http://link.springer.com/chapter/10.1007/978-1-84800-149-7_7.
- Page, Scott E. 2008. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press.
- Rhee, June W., and Eun-mee Kim. 2009. "Deliberation on the Net: Lessons from a Field Experiment." *Online Deliberation: Design, Research, and Practice*, 223–32.

- Schlosberg, D., Shulman, S. W., and Zavestoski, S. 2006. "Virtual Environmental Citizenship: Web-Based Public Participation in Rulemaking in the United States." In *Environmental Citizenship: Getting from Here to There*, 207–36. Cambridge, MA: MIT Press.
- Schneider, Jodi, Tudor Groza, and Alexandre Passant. 2013. "A Review of Argumentation for the Social Semantic Web." *Semantic Web* 4 (2): 159–218.
- Shum, Simon J. Buckingham, Albert M. Selvin, Maarten Sierhuis, Jeff Conklin, Charles B. Haley, and Bashar Nuseibeh. 2006. "Hypermedia Support for Argumentation-Based Rationale." In *Rationale Management in Software Engineering*, 111–32. Springer.
http://link.springer.com/content/pdf/10.1007/978-3-540-30998-7_5.pdf.
- Sia, Choon-Ling, Bernard C. Y. Tan, and Kwok-Kee Wei. 2002. "Group Polarization and Computer-Mediated Communication: Effects of Communication Cues, Social Presence, and Anonymity." *Information Systems Research* 13 (1): 70–90. doi:10.1287/isre.13.1.70.92.
- Sintomer, Yves, Carsten Herberg, Giovanni Allegretti, Anja Röcke, and Mariana Alves. 2013. "Participatory Budgeting Worldwide – Updated Version." *GLOBAL CIVIC ENGAGEMENT – Service for Development Initiatives*.
- Smith, Graham. 2009. *Democratic Innovations: Designing Institutions for Citizen Participation*. 1 edition. Cambridge, UK ; New York: Cambridge University Press.
- Smith, Graham, Peter John, and Patrick Sturgis. 2013. "Taking Political Engagement Online: An Experimental Analysis of Asynchronous Discussion Forums: Taking Political Engagement Online." *Political Studies* 61 (4): 709–30. doi:10.1111/j.1467-9248.2012.00989.x.
- Spada, Paolo, and Giovanni Allegretti. 2014. "The Role of Redundancy and Diversification In Multi-Channel Democratic Innovations." In *Experiments in Democracy – Iceland in Crisis and Recovery*, edited by Jón Ólafsson. Reykjavík: University of Iceland Press & Bifröst University.
- Steenbergen, Marco R, André Bächtiger, Markus Spörndli, and Jürg Steiner. 2003. "Measuring Political Deliberation: A Discourse Quality Index." *Comparative European Politics* 1 (1): 21–48. doi:10.1057/palgrave.cep.6110002.
- Sunstein, Cass R. 2006. *Infotopia: How Many Minds Produce Knowledge*. Oxford University Press, USA.
- Surowiecki, James. 2005. *The Wisdom of Crowds*. Knopf Doubleday Publishing Group.
- Suthers, Daniel D, Ravi Vatrappu, Richard Medina, Samuel Joseph, and Nathan Dwyer. 2008. "Beyond Threaded Discussion: Representational Guidance in Asynchronous Collaborative Learning Environments." *Computers & Education* 50 (4): 1103–27. doi:10.1016/j.compedu.2006.10.007.
- Tapscott, Don, and Anthony D. Williams. 2008. *Wikinomics: How Mass Collaboration Changes Everything*. Penguin.
- Toth, Eva Erdosne, Daniel D. Suthers, and Alan M. Lesgold. 2002. "Mapping to Know?: The Effects of Representational Guidance and Reflective Assessment on Scientific Inquiry." *Science Education* 86 (2): 264–86. doi:10.1002/sce.10004.
- Turner, Eric. 2013. "The 5 Star Movement and Its Discontents: A Tale of Blogging, Comedy, Electoral Success and Tensions." *Interface: A Journal on Social Movements* 5 (2).
<http://www.interfacejournal.net/wordpress/wp-content/uploads/2013/11/Interface-5-2-Turner.pdf>.
- Twardy, Charles. 2004. "Argument Maps Improve Critical Thinking." *Teaching Philosophy* 27 (2): 95–116.
- Van Gelder, T. 2007. "The Rationale for RationaleTM." *Law, Probability and Risk* 6 (1-4): 23–42. doi:10.1093/lpr/mgm032.
- Walton, Douglas. 2006. *Fundamentals of Critical Argumentation*. Cambridge University Press.
- Warren, Mark E., and Hilary Pearse. 2008. *Designing Deliberative Democracy: The British Columbia Citizens' Assembly*. Cambridge University Press.