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

Paolo Spada[1], Mark Klein[2], Raffaele Calabretta[3], Luca Iandoli[4], Ivana Quinto[5], and all the members of the online community of Insieme per il PD.

Abstract[6]

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 proposed 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] University of Southampton
[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? With the advent of the internet many hoped that asynchronous text based discussions[7] could generate new venues of high quality deliberation that could easily scale up (Dahlberg 2001, Dahlgreen 2005). But, as the literature on the quality of online discussions has well documented, these new spaces rarely fulfill such promises (Sunstein 2006; Hartz-Karp and Sullivan 2014, Gervais 2015).

Recent developments in crowdsourcing and idea-sharing software have harnessed the potential of large scale collaborations (e.g.; Wikipedia, Stackoverflow), but have not yet solved the problem of online text based discussions (Lev-on and Hardin 2008). 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 do so without reducing engagement? 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 experiment involved four groups of around 100 citizens each. If we can achieve good e-deliberation in groups containing 100 participants then the possibility of scaling-up deliberation to an actual crowd of thousands of people appears less remote.

[7] 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.

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).[8] This democratic innovation tranfers online a participatory and deliberative decision making process (Intra-party Deliberative Referendums) designed to deepen parties' internal democracy.

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. The Deliberative Referendum embedded a pilot experimental analysis to investigate the effect of different structures of large group e-deliberation. Two randomly generated groups of 160 participants were assigned to discuss using the Deliberatorium, an argument mapping software (Klein, Cioffi, and Malone 2007; Schneider, Groza, and Passant 2013), while two randomly generated groups of 160 participants were assigned to discuss using a standard forum.[9]

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. If argument mapping is too difficult and complex to engage a majority of participants then it cannot be a good technology to scale-up deliberation.

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

[8] http://insiemeperilpd.ning.com/

[9] We employed block randomization to balance the composition of the groups. While 160 people were encouraged to participated, only around 100 people actually did.

mechanisms. If the argument map is written mostly by the moderators it cannot be a good technology to scale up deliberation.

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 an 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; Gutmann and Thompson 2009). The reasoned exchange of arguments is central to virtually all theories of deliberative democracy and it is often identified as the core element of a minimal conception of deliberation (Landemore 2013; Mansbridge 2015). The characterization of what constitutes a reasoned argument in the literature has evolved over time, from a more rational and restrictive paradigm, to a more encompassing conception that includes claims based on experiential knowledge, emotions and story-telling (Bächtiger et al. 2010). There are many other important facets of the quality of deliberation (Young 2002; Gutmann and Thompson 2009; Bächtiger et al. 2010), but, ceteris paribus, if we can increase the amount of reasoned argumentation in a group discussion the overall quality of deliberation will improve.

How well do existing online technologies meet the challenge of promoting 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, brainstorming and participation at a vast scale, which in turn can lead to remarkably powerful emergent collaborative phenomena (Lev-On, Azi and Hardin, Russell 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 and argumentation online has not really changed in the past thirty years and is based on asynchronous text based interactions. These large group e-discussions are prone to dysfunctional emergent behaviors[10] that can deeply undercut the quality of the deliberation outcomes. Problems such as stronger group polarization (Sunstein 2006; Chen 2013), informational cascades (Hansen, Hendricks, and Rendsvig 2013), low signal-to-noise ratio (Cotton and Yorke 2006), information overload (Losee 1989), scattered content, redundancy, non-collaborativness plague asynchronous text based interactions.

Overall a variety of pilot projects have shown that it is possible to elicit massive collaboration in the brainstorming phase[11] or in the voting phase of an online participatory process[12], what remains very difficult to achieve is to elicit large scale high quality argumentation (Klein and Convertino 2014).

[10] 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.
[11] Open innovations systems are a clear example of this possibility. For example the google project 10 to the 100 managed to elicit in a few months 150000 ideas in 2008 (http://www.google.com/campaigns/project10tothe100/).
[12] For example the e-voting platform of the participatory budgeting process of the state of Rio Grande do Sul every year attracts more than 170000 voters Spada et al. 2016.

Due to the widespread use of forum and chat 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.[13] 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 quantity and quality of argumentation 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-voting[14], provision of information, geo-mapping) without promoting significant horizontal interactions among participants (Schlosberg, D., Shulman, S. W., and Zavestoski, S. 2006). This concatenation of events generates a vicious cycle that reinforces the existing beliefs of the low quality of online deliberations.

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.

3. LARGE SCALE ARGUMENTATION

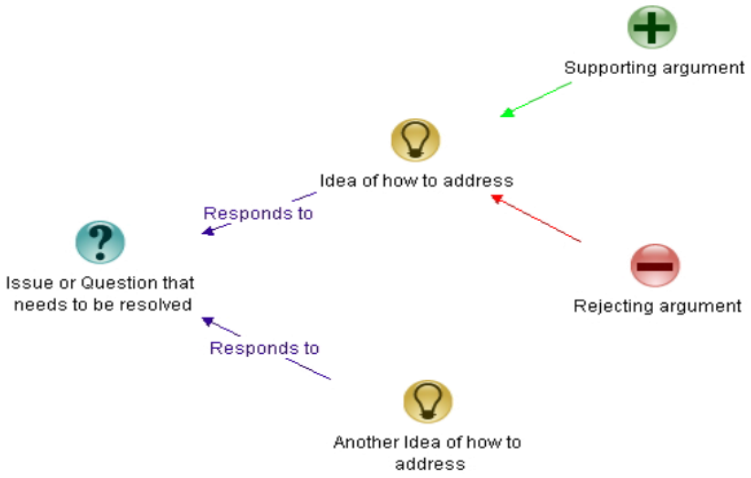
Large-scale argumentation represents a promising approach to addressing some of the weaknesses of current deliberation technologies. Argumentation tools (Schneider, Groza, and

[13] Godwin's law is an Internet adage asserting that "As an online discussion grows longer, the probability of a comparison involving Nazis or Hitler approaches 1" https://en.wikipedia.org/wiki/Godwin%27s_law

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

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



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 (Buckingham-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 Issue-Based Information System (IBIS) argumentation formalism (Conklin 2006), which is one of the simplest and most widely applied. The elements of an IBIS argument map are issues (or questions that need to be answered), ideas (or solutions to the issues) and arguments which support or object to a given idea (or another argument).

In the Deliberatorium the moderators' role 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.

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

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 this 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 evryday life when a party implements a Deliberative Referendum.[15] Our prior was that the Deliberatorium might be so costly for party members without an engineering degree that the advantages shown in pilot project with students would disappear. Retention rates would plummet and participation would be minimal. As we will see we were completely wrong. But before presenting the results of the experiment, in the next section we sketch out the other fundamental elements of the democratic innovation we designed.

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

The Intra-party Deliberative Referendums are democratic innovations aimed at improving political parties' internal decision making. Three key elements distinguish the Deliberative Referendum from traditional discussion within parties.

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

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 Referendums were just a proof of concept discussed within the parties and civil society organization of the Italian left. Therefore, for this experiment we designed an Intra-party Deliberative Referendum drawing on best practice from democratic innovations.

Similarly to Citizens' Assemblies and Reference Panels 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 quasi-representative sample of the population. Using the enrollment survey we created four randomly generated discussion groups of 160 participants each.[16] 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 used block randomization to balance the four discussion groups across gender, age, education, intensity of preferences over the two main families of electoral laws (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

[16] We encouraged 160 citizens to participate, and around 100 did in each group.

discussions.[17] 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.[18] Given that one of the common problems of participatory budgeting is the mismanagement of expectations, 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 the various opportunities 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 email as long as the information was purely descriptive and did not contain any persuasive message.[19]

[17] We used block randomization to improve both efficiency and balance of observed and unobserved covariates across treatment conditions. We checked the success of the randomization with respect to observables by fitting multinomial logit models. The block randomization procedure was successful in balancing these observables. The results are available upon request.

[18] 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.

[19] An Italian mathematician sent his custom electoral system through this method.

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. 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 per enrolled user.[20] A team of nine moderators monitored the discussions in all groups 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 three weeks our moderators did not have to ban a single participant for disruptive behavior or inappropriate language.

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

[20] In the enrollment phase the ip-address and the email were automatically checked and a subsequent manual screening was conducted.

and two groups of 160 participants to discuss using the Deliberatorium. Each discussion lasted three weeks. Using a pre-survey and block randomization, 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 (see online appendix) 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 emails, every 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 next intra-party deliberative referendum?” Users could make whatever contributions they wanted within these fixed macro-issues.

The e-deliberation ran for three weeks, from April 1 to April 21, 2012. After that we constructed a ballot and we allowed users to vote for the ideas that the party should support. Users were also given a brief post-survey that asked about their experiences with the systems. Around fifty percent of the enrolled participants were active in the discussion and about three-quarters of the active participants voted and filled out the post survey. Around ten percent of the enrolled users was not active in the discussion, but participated in the final vote. 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 classic methods of comparison frequently employed by the literature on deliberative democracy, such as civility of discussion, integrative complexity (Brundidge et al. 2014), the Deliberation Quality Index (Steenbergen et al. 2003), or measures of noise, would automatically generate better results for the map or the forum by construction. For example, 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 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 the content of the 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.

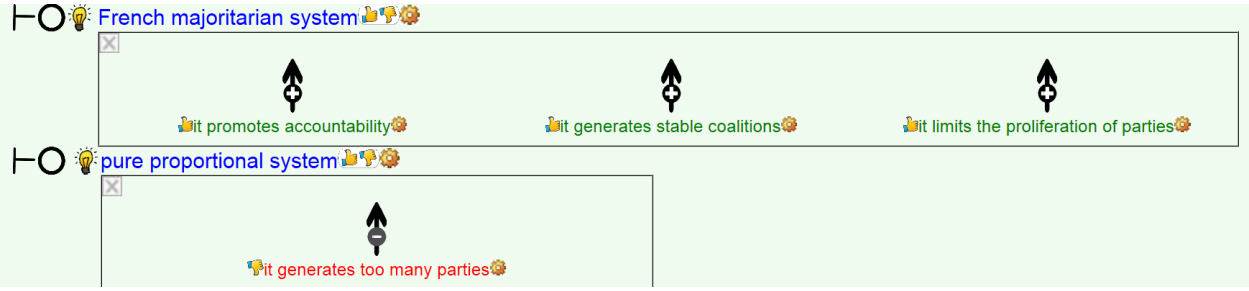
Each post in the map has to be certified by a moderator and has to be justified. Noise, e.g., off topic posts, is not certified and thus is not visible to users.[21] The map does not offer the space for non-constructive or disrespectful speech acts. Thus indexes that measure noise and civility of the 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).

[21] 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 metaposts 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.

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.

Figure 2: Map representation



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. Thus storytelling and type 2 deliberation, i.e. deliberation that is less rational and more emotional, is included in the discussion and not discarded (Bächtiger et al. 2010).

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.

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.[22] 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. For example Google's project10tothe100.com generated 150,000 ideas and the company had to hire 3000 employees to filter them, a process that took more than nine months. Lisbon's participatory budgeting process has experienced an identical problem. To face the almost 1,000 proposals of investments generated by citizens every year in face-to-face and online meetings, the municipality has organized an Interdisciplinary Working Group of Civil Servants to merge and pre-select the proposals. Every year since 2009 the list of around 200 filtered projects sparks numerous complaints by citizens who see their ideas disappearing or being distorted (Sintomer & Allegretti, 2016). We had only around a thousand forum posts to evaluate, thus the process was less complex, but it still required more than 160 hours of work.

[22] 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.

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 number of ideas that received at least one comment, the depth of the map and the depth of the ideas and sub-ideas, and the depth of the arguments. These measures are meant to explore which tool is more capable of gathering ideas and exploring issues and arguments in favor of or against an idea, in other words our measure focus on the argumentation aspect of deliberative performance (Mercier and Landemore 2012).

6. OUTCOMES

Of the 639 enrolled participants, 356 (55%) 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[23], but if the deliberation had continued participation probably would have risen further (Figures 3 and 4).

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. Figures 3 and 4 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 these processes and when participation plateaus on the basis of different groups' size, intensity of advertising and type of deliberation.

[23] The experimental design was approved by Harvard University IRB process.

Figure 3: Cumulative number of first login per day Forums

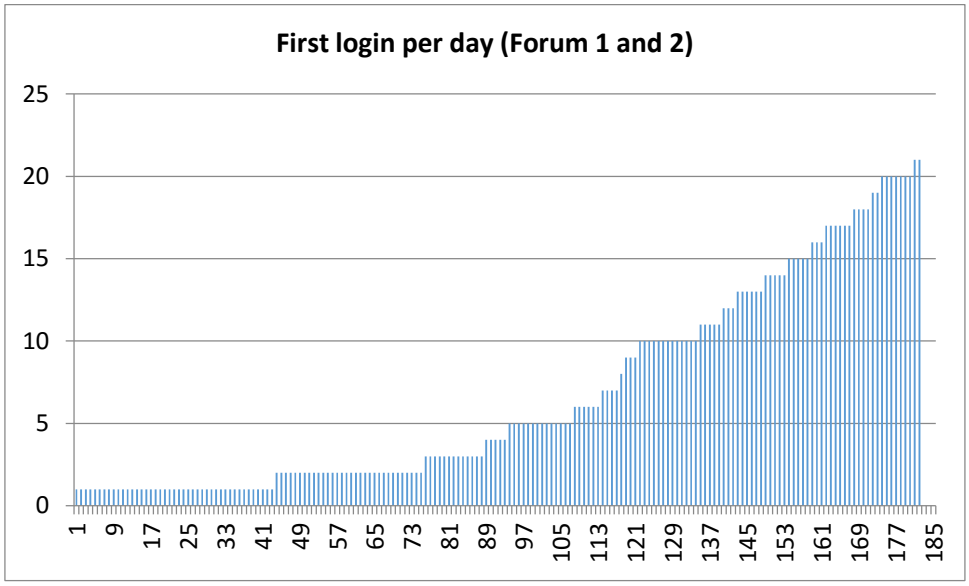
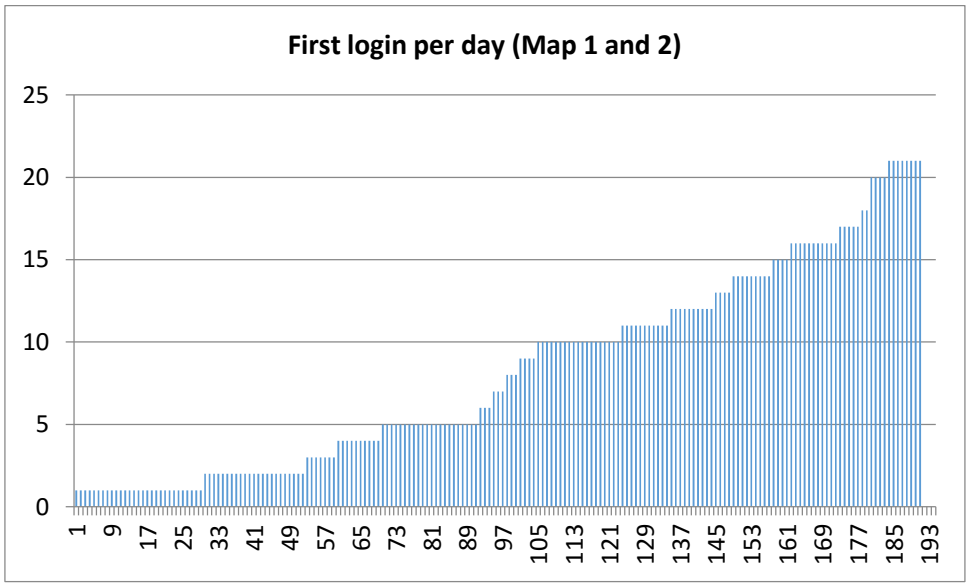


Figure 4: Cumulative number of first login per day Argument Maps



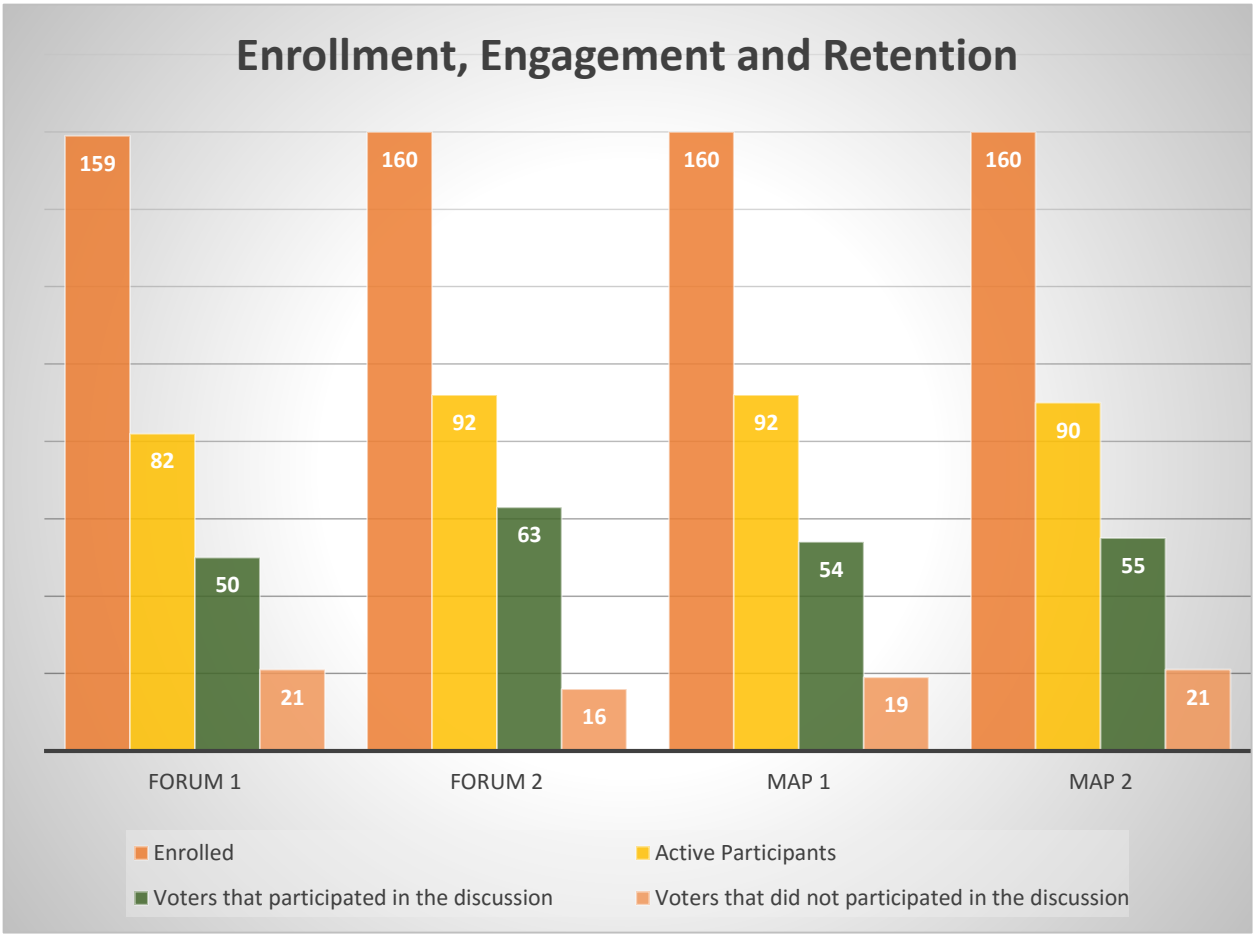
We had minimal resources for advertising, thus these graphs likely represent a lower bound of what can be achieved in similar processes within parties.

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). For such reasons in this study we take the most conservative route and we analyze the effect of the treatment on group level measures of argumentation restricting our sample to N=4. In a separate study we instead analyze the impact of the software on individual attitudes using a pre and post survey design (Authors 2016).

Figure 5 shows that there was no difference with respect to engagement (number of people that participated in the discussion), and retention (number of people that completed the discussion and cast a vote) among the various treatment groups. In each treatment group more than 50% of people were engaged, and around 60% kept participating till the end. The two treatments generated similar average number of participant per day, and similar retention rate.[24] While engagement was mainly a function of the advertising campaign, retention rates were significantly influenced by the treatments. Most surprisingly the implementation of the Deliberatorium has not negatively impacted retention rates. This is a counterintuitive result, given the complexity of participating in the co-generation of an argument map.

[24] The average number of participants per day was 11.2 in forum 1, 13 in forum 2, 13.5 in map 1, 13.8 in map 2.

Figure 5: Enrollment, Engagement and Retention



When we investigate the other major event type in figure 6 (View a post, Create a post, Contribute to a certified post, Rate a post, View the virtual library) we observe statistically significant differences only in the rate a post category. The argument map increases the percentage of users that rate a post by 23 percentage points.[25] As expected our advertising campaign had no impact on the number of people that participated in each treatment group, differences are due to idiosyncratic characteristics of each group. On average the Forum

[25] Using a two-tail t-test for the difference in mean the treatment effect is significant at 1% level. The 99% confidence interval is 14%, 33%.

conditions have slightly more users that created a post (+4%), but this difference is not significantly different from zero probably due to sample size.

Figure 6: number of unique users who performed each major event type at least once.

Event Type	Forum 1	Forum 2	Map 1	Map 2
Viewed a post	70 (44%)	82 (51%)	76 (48%)	76 (48%)
Created a post	35 (22%)	42 (26%)	32 (20%)	30 (19%)
Rated a post	15 (9%)	18 (11%)	53 (33%)	54 (34%)
Entered the virtual library	25 (16%)	30 (19%)	29 (18%)	31 (19%)
Enrolled in the experiment	159 (100%)	160 (100%)	160 (100%)	160 (100%)

We hypothesize that rating was more frequent in maps than forums because, in the forum, ratings are less meaningful; a forum post may make several points, so a rating 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.

When we look at the overall activity in the four treatment groups we see that 139 participants, or approximately 22% 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 (304), we see that almost half of them created content (46%). Given the 1/9/90 rule of thumb of internet participation these results are impressive and exceed our expectations.[26] This unusual level of participation might have been

[26] 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.

generated by a combination of factors including the controversial topic that was concurrently being covered by the media[27], the design of the Intra-party Deliberative Referendum that targets the supporters of a political party that have a shared value system and are interested in politics, the self-selected sample, the design of the process itself, and by a novelty effect.

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 in that it has no significant impact on retention or activity levels. This is a completely unexpected 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. All the details of the experiment were explained to the users before its beginning. 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 person-hours. The moderation of the forum was minimal because the discussion was extremely civil and required ten person-hours which mostly involved reading posts. The moderation of the

[27] 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.

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 facilitated by the help chatline which offered numerous examples of common errors and mistakes. In total we spent 42 person-hours for the map conditions. Overall the traditional e-consultation method, forum plus expert post-processing, required four times more person-hours than the maps (170 vs 42). However, it is important to keep in mind that the moderators of an argument mapping software require more training than the moderators of a traditional forum, and that in many online consultations the postprocessing does not require the construction of a full argument map. Our calculation cannot reflect these two factors, because to prevent bias we employed the same set of moderators in the two treatments, and we cannot provide an estimate of a real-life postprocessing because a standard procedure does not exist. On the other hand, our experiment was relatively small, in larger projects with hundred of thousands of contributions the amount of noise and redundancies present in the forums would increase further requiring an exponential increase in postprocessing time. In sum these estimates have limited external validity and further research is required, but they offer a first useful proxy of the efficiency gains that can be obtained by using argument mapping.

Comparing the self-generated map with the forum post-processed by a team of argument mapping experts (Figure 7), we observe that the forums in total created more than three time as many ideas as the user-generated map (290 vs 82).

Figure 7: Group level raw data, comparing the two treatments[28]

Group (N=4)	Ideas	Pros	Cons	Issues	Ideas with one argument	Ideas with one argument by others	Ideas with no argument
Forum 1 & experts	155	92	85	16	50	35	105
Forum 2 & experts	135	53	43	3	27	22	108
Map 1	31	33	32	7	18	16	13
Map 2	51	39	59	10	21	21	30

On average the map treatment generates 104 less ideas than the forum treatment. This average treatment effect is significantly different from zero in a two-tailed test (p=0.013, 95% conf. interval -165;-43) even if the unit of analysis is the group and thus the sample size is 4.

While the effect on the map on idea generation 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. This result is driven by the proliferation of ideas that are not discussed (Ideas with no argument) in the forum. The map treatment generates on average 85 less ideas with no argument than the forum treatment. The latter result is significantly different from zero in a two-tailed test (p=0.01, 95% conf. interval 48;122). The map treatment instead has no significant negative effects on the generation of ideas that generates discussion (Ideas with one argument; Ideas with one argument by others).

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 again not significant. In order to capture in

[28] 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.

relative terms the effect of the map treatment, we investigate the following argumentation indexes:

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

$$Argumentation\ index\ 2 = \frac{ideas\ that\ elicited\ argumentation\ from\ others}{ideas}$$

The first index captures the amount of argumentation per idea in each group, the second index capture the ratio of ideas that elicited argumentations from other authors. The basic intuition is that we want to normalize the amount of arguments exchanged horizontally across users by the overall level of activity and by the overall chances of providing arguments. The more ideas are present in an environment the easier is for a user to find one that they can provide a comment to.

Figure 8: Group level raw data, comparing metrics of argumentation

Group (N=4)	Arg. Index 1	Arg. Index 2
Forum 1 & experts	1.24	0.23
Forum 2 & experts	0.73	0.16
Map 1	2.32	0.52
Map 2	2.11	0.41

The difference in the average effect of the map and the forum on index 1 is 123 percentage points, i.e., at the group level the self-generated map creates more than twice as many arguments per idea than the forum post-processed by the experts; this result is significantly different from zero in a two-tailed test (p=0.041; 95% conf. interval 0.04; 2.41). The difference in the average effect of the map and the forum on index 2 is smaller, 27 percentage points, i.e. the map

increases the percentage of ideas that elicit at least one comment by another user by a quarter, this result is significantly different from zero in a two-tailed test (p=0.047; 95% conf. interval 0.01; 0.53).

The two argumentation indexes highlight a key difference between the two treatments: the map promotes a richer exchange of arguments per idea than the forum. The small number of groups employed by this pilot recommend caution, but clearly points in the direction of potentially large effects of this e-deliberation technology.

Figure 9: Group topology

Group	Avg. depth[29]	Std. Dev. Depth	Avg. breadth[30]	Std. Dev. breadth	Avg. Adepth[31]	Std. Dev. 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

Not surprisingly when we look at the average depth and average breadth (Figure 9) 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 from the raw data of the forum shows a significant increase in depth of 0.32 (p=0.01, 95% conf. interval 0.19; 0.46) and a significant increase in the

[29] 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.

[30] 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).

[31] The depth of the arguments of the map (Adepth) captures the depth of the chain of pros and cons assuming that the first pro or con has depth zero independently of its actual depth.

breadth of 1.03 (p=0.02, 95% conf. interval 0.41; 1.66). 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. Users in the forum tended to propose a lot of small variations of family of ideas that had to be reorganized by experts in many families and sub-families of ideas. For example a typical user would propose the idea bundle A+B+C (e.g. Majoritarian system, with single member district, and runoff), and the subsequent users would propose and discuss a number of variations and expansions on such idea (e.g., gender quota, small districts, mandatory primaries, campaign finance reform). Thus the post-coding process had not only to split each bundle in its core components, but also create families of ideas that had common idea parts.

However when we look at the depth of the arguments (Adepth), 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. 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 just 4 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 a first step toward the concrete possibility of inducing good argumentation in large groups of 100 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 careful design of the engagement campaign of the Intra-party Deliberative Referendum contributed critically to the quality of the e-deliberation. The Intra-party Deliberative Referendum is a multi-channel system of engagement that was designed with reference to the most successful implementations of face to face democratic innovations and the experience of the largest online community of the Italian Democratic Party with social network advertising. Our experiment is not designed to separate the effect of the Deliberatorium from the effect of the Intra-party Deliberative Referendum. What our experiment shows is that we can use the Deliberatorium to improve a carefully designed e-deliberation process by inducing better argumentation than traditional threaded forum technology. Compared to the forum, the Deliberatorium increases the number of arguments per idea in each group by 123 percentage points and increases the number of ratings per contribution by 23 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 synthetized 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. Through a dialogue with moderators 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. Such meta-discussions are successfully implemented by many participatory budgeting programs around the world to design and update the internals rules of such democratic innovations.

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 (100 participants), that would imply, at least, a 10-fold expansion of the size of face to face small-groups adopted in current deliberations. Current participatory planning processes routinely integrate 10 small groups discussion (50 to 100 participants), applying the same integration mechanism using the Deliberatorium would generate deliberations that include 1000 people.

It is also important to note that the results we are presenting are conditional to the sample of self-selected participants (see the demographics section of the online appendix). 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 Intra-Party 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, very preliminary, step in optimizing an online environment that is capable of sustaining good deliberation involving thousands of people.

REFERENCES

Bächtiger, A., Niemeyer, S., Neblo, M., Steenbergen, M. R., and Steiner, J. (2010). Disentangling Diversity in Deliberative Democracy: Competing Theories, Their Blind Spots and Complementarities*. *Journal of Political Philosophy* 18, no. 1: 32–63.

Black, L. W., Burkhalter, S., Gastil, J. and J. Stromer-Galley. (2010). Methods for Analyzing and Measuring Group Deliberation. In E. P. Bucy and R. L. Holbert, eds. *The Sourcebook for Political Communication Research: Methods, Measures, and Analytic Techniques* New York, NY: Routledge.

Brundidge, J., Reid, S. A., Sujin Choi, and Muddiman, A. (2014). The 'Deliberative Digital Divide:' Opinion Leadership and Integrative Complexity in the U.S. Political Blogosphere: Opinion Leadership in Political Blogs. *Political Psychology*.

Chen, Hsuanwei Michelle. (2013). Group Polarization in Virtual Communities: The Case of Stock Message Boards. At https://www.ideals.illinois.edu/handle/2142/36051, accessed August 19, 2014.

Cohen, J. (1989). Deliberation and Democratic Legitimacy. *Debates in Contemporary Political Philosophy*: 342.

Coleman, S., and Shane, P. M. (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, D., and J. Yorke. (2006). Analysing Online Discussions: What Are Students Learning. *23rd Annual Ascilite Conference: Who's Learning*. At http://www.ascilite.org/conferences/sydney06/proceeding/pdf_papers/p65.pdf, accessed August 20, 2014.

Dahlberg, L. (2001). The Internet and democratic discourse: Exploring the prospects of online deliberative forums extending the public sphere. Information, Communication & Society, 4(4), 615–633.

Dahlgren, P. (2005). The Internet, public spheres, and political communication: Dispersion and deliberation. Political Communication, 22(2), 147–162.

Doan, A., Ramakrishnan, R., and A. Y. Halevy. (2011). Crowdsourcing Systems on the World-Wide Web. *Communications of the ACM* 54, no. 4: 86.

Elster, J. (1998). *Deliberative Democracy*. Cambridge University Press.

Gerber, A. S., and D. P. Green. (2012). *Field Experiments: Design, Analysis, and Interpretation*. W. W. Norton.

Gervais, B. T. (2015) Incivility Online: Affective and Behavioral Reactionsto Uncivil Political Posts in a Web-based Experiment, *Journal of Information Technology & Politics*, 12:2, 167-185.

Gürkan, A., Iandoli, L., Klein, M. and G. Zollo. (2010). Mediating Debate through on-Line Large-Scale Argumentation: Evidence from the Field. *Information Sciences* 180, no. 19: 3686–3702.

Gutmann, A., and D. Thompson. (2009). *Why Deliberative Democracy?*. Princeton University Press.

Habermas, J. (1984). *The Theory of Communicative Action: Reason and the Rationalization of Society*. Beacon Press.

Hansen, P. G., Hendricks, V. F. and R. K. Rendsvig. (2013). Infostorms. *Metaphilosophy* 44, no. 3: 301–26.

Hartz-Karp, J., and B. Sullivan. (2014). The Unfulfilled Promise of Online Deliberation. *Journal of Public Deliberation* 10, no. 1: 16.

Isernia, P., and J. S. Fishkin. (2014). The EuroPolis Deliberative Poll. *European Union Politics* 15, no. 3: 311–27.

Karpowitz, C. F., Mendelberg, T. and L. Shaker. (2012). Gender Inequality in Deliberative Participation. *American Political Science Review* 106, no. 03: 533–47.

Kirschner, P. A., Buckingham-Shum, S. J. and C. S. Carr. (2003). *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*. Springer Science & Business Media.

Klein, M., M. Cioffi, and T. Malone. (2007). Achieving Collective Intelligence via Large-Scale On-Line Argumentation. *ICIW*. At http://cci.mit.edu/publications/CCIwp2007-01.pdf, accessed August 20, 2014.

Klein, M., and G. Convertino. (2014). An Embarrassment of Riches. *Communications of the ACM* 57, no. 11: 40–42.

Landemore, H. (2013). On Minimal Deliberation, Partisan Activism, and Teaching People How to Disagree. *Critical Review* 25, no. 2: 210–25.

Lev-On, Azi and Hardin, Russell (2008) Internet-Based Collaborations and Their Political Significance, *Journal of Information Technology & Politics*, 4:2, 5-27.

Losee, R. M. 1989. Minimizing Information Overload: The Ranking of Electronic Messages. *Journal of Information Science* 15, no. 3: 179–89.

Mansbridge, J. (2010). Deliberative Polling as the Gold Standard. *The Good Society* 19, no. 1: 55–62.

———. (2015). A Minimalist Definition of Deliberation. *Deliberation and Development: Rethinking the Role of Voice and Collective Action in Unequal Societies* Washington, DC: World Bank.

Mercier, H., and H. Landemore. (2012). Reasoning Is for Arguing: Understanding the Successes and Failures of Deliberation: Reasoning Is for Arguing. *Political Psychology* 33, no. 2: 243–58.

Okada, A. (2008). Scaffolding School Pupils' Scientific Argumentation with Evidence-Based Dialogue Maps. *Knowledge Cartography* Springer. At http://link.springer.com/chapter/10.1007/978-1-84800-149-7_7, accessed September 11, 2014.

Page, S. E. (2008). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press.

Schlosberg, D., Shulman, S. W., and S. Zavestoski (2006). Virtual Environmental Citizenship: Web-Based Public Participation in Rulemaking in the United States. *Environmental Citizenship: Getting from Here to There* Cambridge, MA: MIT Press.

Schneider, J., Groza, T. and A. Passant. (2013). A Review of Argumentation for the Social Semantic Web. *Semantic Web* 4, no. 2: 159–218.

Sintomer, Y. and Allegretti, G. (2016, forthcoming). Os Orçamentos Participativos na Europa. Entre democracia participativa e modernização dos serviços públicos. Almedina, Coimbra.

Buckingham-Shum, S. J., Selvin, A. M., Sierhuis, M., Conklin, J., Haley, C. B., and B. Nuseibeh. (2006). Hypermedia Support for Argumentation-Based Rationale. *Rationale*

Management in Software Engineering Springer. At http://link.springer.com/content/pdf/10.1007/978-3-540-30998-7_5.pdf, accessed August 20, 2014.

Smith, G., John, P. and P. Sturgis. (2013). Taking Political Engagement Online: An Experimental Analysis of Asynchronous Discussion Forums: Taking Political Engagement Online. *Political Studies* 61, no. 4: 709–30.

Spada, P., Mellon, J., Peixoto, T. and F. M. Sjoberg. (2016). Effects of the Internet on Participation: Study of a Public Policy Referendum in Brazil. *Journal of Information Technology & Politics*.

Steenbergen, M. R, Bächtiger, A., Spörndli, M. and J. Steiner. (2003). Measuring Political Deliberation: A Discourse Quality Index. *Comparative European Politics* 1, no. 1: 21–48.

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

Suthers, D. D., Vatrapu, R., Medina, R., Joseph, S. and N. Dwyer. (2008). Beyond Threaded Discussion: Representational Guidance in Asynchronous Collaborative Learning Environments. *Computers & Education* 50, no. 4: 1103–27.

Van Gelder, T. (2007). The Rationale for Rationale™. *Law, Probability and Risk* 6, no. 1-4: 23–42.

Warren, M. E., and H. Pearse. (2008). *Designing Deliberative Democracy: The British Columbia Citizens' Assembly*. Cambridge University Press.

Young, I. M. (2002). *Inclusion and Democracy*. Oxford University Press.