

Supporting Reflective Public Thought with ConsiderIt

Travis Kriplean¹, Jonathan Morgan², Deen Freelon³, Alan Borning¹, Lance Bennett^{3,4}

Computer Science¹, Human Centered Design & Eng.², Communication³, Pol. Science⁴
University of Washington, Seattle

ABSTRACT

We present a novel platform for supporting public deliberation on difficult decisions. ConsiderIt guides people to reflect on tradeoffs and the perspectives of others by framing interactions around pro/con points that participants create, adopt, and share. ConsiderIt surfaces the most salient pros and cons overall, while also enabling users to drill down into the key points for different groups. We deployed ConsiderIt in a contentious U.S. state election, inviting residents to deliberate on nine ballot measures. We discuss ConsiderIt's affordances and limitations, enriched with empirical data from this deployment. We show that users often engaged in normatively desirable activities, such as crafting positions that recognize both pros and cons, as well as points written by people who do not agree with them.

Author Keywords

governance, deliberation, consideration, politics, reflection

ACM Classification Keywords

H.5.0 Information Interfaces and Presentation: General

INTRODUCTION

My Way is to divide half a Sheet of Paper into two Columns, writing over the one Pro, and over the other Con. Then during three or four Days Consideration I put down under the different Heads short Hints of the different Motives that at different Times occur to me for or against the Measure...I find at length where the Ballance lies...And tho' the Weight of Reasons cannot be taken with the Precision of Algebraic Quantities, yet when each is thus considered separately and comparatively, and the whole lies before me, I think I can judge better, and am less likely to take a rash Step. – Ben Franklin, inventor of modern pro/con list [5]

Franklin recognized the value of challenging himself to consider tradeoffs through *personal deliberation*. For decisions that involve many people, engaging others in *public deliberation* has further benefits, such as making better decisions, uncovering new solutions, and including diverse others in decisions [13, 20, 19]. Yet public deliberation is difficult. It is challenging for people to listen to perspectives that contradict their own; most people, at least in America, are averse to

even engaging with those with whom they suspect of harboring different views [36]. The use of communication media reflect these tendencies, with people highly attracted to content that aligns with their prior beliefs, even those who do not actively avoid challenging content [18, 27, 28, 35]. The problem is reinforced by a lack of effective socialization in the skills and values associated with productive exchanges of ideas, particularly in public schools [38]. We find our capacity to publicly deliberate undercut by people framing issues as zero-sum games between opposing sides, with members of each side speaking past each other and missing opportunities to factor other perspectives into their thoughts.

We think we can do better. By building interfaces that subtly encourage people to consider issues and reflect on the diverse perspectives of others, we think we can help build public trust while improving upon our collective ability to take more effective action on problems such as financial reform and climate change. We offer a modest step in this direction with the ConsiderIt platform.

ConsiderIt builds from the basics of personal deliberation to foster more effective public deliberation. It focuses people on thinking through the tradeoffs of a proposed action, such as a ballot measure in an election, by inviting them to create a pro/con list. ConsiderIt augments this familiar activity by enabling users to not just author pro and con points, but to *include* into their own lists the points others have already contributed, and in turn share with others the points that they author. The considerations of others thus become raw material for one's own considerations. Complementing the pro/con lists, participants summarize their stance on an issue on a continuum of conviction, rather than a binary yes/no vote. The focus is on augmented personal deliberation, rather than direct discussion with others, thus potentially mitigating the activation of political identity and flaming.

ConsiderIt then repurposes these personal deliberations to offer an evolving guide to public thought. Here, ConsiderIt surfaces the most salient overall pro/con considerations based on how often they are included and whether they are included by people with different stances on the issue. ConsiderIt also enables drill down into the salient points for different segments of the population: “*What were those who strongly opposed this thinking???*” Users can thus gain insights into the considerations of people with different perspectives, rather than making assumptions based on caricatures. As we describe in our results, this can help users identify unexpected common ground. We also contribute a pro/con ranking metric tailored to highlight points that resonate with a diverse public, to promote persuasive points while still encouraging a diversity of views, and to hopefully be resistant to strategic manipulation.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CSCW'12, February 11–15, 2012, Seattle, Washington, USA.

Copyright 2012 ACM 978-1-4503-1086-4/12/02...\$10.00.

ConsiderIt is the result of a design process initiated with the high-level goal of supporting a community-written voters guide for Washington State in the contentious 2010 U.S. election. In this paper, we ground our system contribution by reporting on ConsiderIt's deployment in September 2010 as *The Living Voters Guide* (LVG). It attracted thousands of users and hundreds of contributors in the five weeks leading up to the election. This style of pragmatic inquiry reflects our philosophy for making progress on the broader, complex challenges of civic participation: deploy a novel system into a real situation and examine whether it has basic utility for people using it at their own discretion. From this vantage point, we can then describe the approach's particular affordances and limitations as they become apparent through use.

This paper thus presents generative social system research: (1) we have *created a new interactive approach* for supporting public deliberation that has successfully weathered a large field deployment; (2) we describe a design strategy that *may spark new ideas*; and (3) the deployment *identifies opportunities for future controlled experimentation*. Our field deployment complements hypothesis testing in a lab, gaining insight into the utility and affordances of a social system operating in a contested and often vitriolic domain. After discussing past approaches to supporting public deliberation and the theoretical underpinning of this work, we present ConsiderIt in detail. We then describe the LVG deployment, followed by an analysis of ConsiderIt's appropriation. We end with a discussion of challenges and future directions.

THEORY AND DESIGN FOR PUBLIC DELIBERATION

Any organized group must make tough decisions when setting policy and choosing how to allocate scarce resources. Traditionally, CSCW researchers have focused on small-group decision-making embedded in hierarchical organizations [29]. However, with the advent of social media, it is now possible to assemble larger publics to help deliberate, whether these constituents are citizens, customers, a student body or employees. The input can be merely consultative, as with Facebook's Site Governance initiative for gathering feedback on terms of service changes or U.S. President Obama's use of platforms for gathering ideas about implementing open governance. Or they can run deep in the DNA of the organization, such as in Wikipedia where contributors struggle to conduct distributed self-governance [6, 14, 21]. As we move outside hierarchical organizations, we believe we should expand our guiding theory to incorporate insights from political communication in order to more effectively design for these wider, more casual engagement efforts.

Deliberation is just one normative ideal of collective decision-making, and it is sometimes in conflict with others. Social scientists have begun to draw on political theory to juxtapose and measure these conflicting ideals [2, 8, 16, 48]. Freelon [16] maps out three such ideals and their associated behavioral indicators within online spaces. Aside from the deliberative style, liberal individualism [10] is an egocentric perspective celebrating self-expression over listening, manifesting in monologues and flaming. Liberal individualist behavior is prevalent online, occurring primarily in spaces that are ideologically heterogeneous, such as the free-for-all comment sections of Youtube and major newspapers, which both

offer broad freedom of expression but are abrasive and low in responsiveness [31, 47]. Communitarianism [15, 46] emphasizes group identification with like-minded people in order to better organize against groups with whom they disagree. Evidence from studies of ideological fragmentation and selective exposure have demonstrated that communitarian behavior is exceedingly common both online and off, with most people choosing to engage material that aligns with their prior beliefs [18, 27, 28, 35]. And while some people do not go out of their way to avoid challenging material, they still must encounter it first, an activity that platforms like Facebook and Google Reader do not encourage.

While we respect the importance of these different ideals, most, if not all, broadly used communication interfaces implicitly or explicitly support these non-deliberative styles. Some scholars and practitioners have developed methods and tools in response to this deficit in deliberative support. However, these efforts have yielded tools and methods requiring a high level of investment from both administrators and users, which limits both their scope and sustainability. Numerous methods have been invented to approximate the deliberative ideal in face to face settings, such as deliberative polling and citizen juries [13]. Most systems work in this domain has tried to translate aspects of these multi-day, resource intensive deliberative efforts into online tools [24, 25, 34]. They tend to centralize control over the agenda and delimit acceptable contributions, such as by instituting Robert's Rules of Order [40, 41]. One particularly well-explored class of high-investment deliberative system disciplines participants to break down their opinions according to argumentation schemas drawn from linguistic theory or elsewhere, and incorporate them into a graphical visualization of the issues, arguments, positions, and so forth (see [43, 44] for an overview). Grave challenges to this approach have been found in practice, such as difficulty in learning the formalized schemas, breaking up narratives into the required fine chunks, and agreeing on classifications [23, 42], often requiring trained facilitators to be successful [7].¹ These systems have their place, but it would be unrealistic to expect them to be used widely. Aside from the unsustainable level of commitment demanded, not everyone shares the deliberative ideal. Insisting on a deliberative style in a discretionary setting may simply lead to low participation.

Our position is that if we want to make progress on addressing collective problems, we should try to support deliberative styles of interaction in more casually used interfaces, and further, in a way that accommodates aspects of the other styles without undermining deliberation. If we can attract the attention of diverse communicators, the interface can gently nudge people toward deliberative behavior, such as weighing trade-offs and considering what other people are saying about the issue. We may thus be able to promote an environment where people are more likely to find common ground and less likely to participate in flame wars. This is certainly a challenge, but Mackuen [33] has shown that the communicative styles that people adopt are context sensitive. And interface design can play a large role in shaping this context [37, 45].

¹Recent work such as Cohere [44], Pathfinder [32] and Videolyzer [9] show promise by relaxing some of the constraints of the argumentation scheme.

For example, Park et al. demonstrated that displaying multiple articles about the same news event, divided into clusters that emphasize different aspects of the event, causes people to read more diverse news stories than a simple random list [37], even though most people appear to desire only news stories that back their own beliefs [35].

Therefore, while it is difficult to design interfaces for wide public use that subtly nudge people toward a deliberative style of communication, we believe it is possible. Our approach to this challenge has three parts. First, we try to encourage people to explicitly consider tradeoffs for the issue at hand. In ConsiderIt, this is primarily facilitated by the pro/con list, where an unbalanced list challenges users to reflect on whether they are missing something. Second, we encourage people to listen to and recognize what others have to say. In ConsiderIt, this is primarily facilitated by enabling users to adopt into their own pro/con lists points contributed by others. When users adopt a point, they are both using it to consider a tradeoff, and also listening to someone else. (Whether they are cognizant of this dual role is another issue.) Third, we aggregate and repurpose these explicit acts of considering tradeoffs and listening to others to help everyone better understand and explore facets of the issue being deliberated. In ConsiderIt, this is facilitated by providing a view over the deliberations that ranks the pros and cons based how often and by whom the points are adopted.

Few other systems are designed with a similar strategy. Reflect [30] modifies the comment sections of webpages to facilitate the establishment of common ground by adding a listening box next to every comment, where other users are encouraged to succinctly restate the points that the commenter is making. This is a nudge to listen to other users. Other readers can then read the original comment and the listeners' interpretations of what was being said, supporting broader understanding of the discussion. OpinionSpace [12] plots on a two-dimensional map the individual comments in a web forum, based on the commenters' responses to a short value-based questionnaire. By navigating this space, readers are better able to seek out a diversity of comments as well as prime themselves for engaging the perspective of someone with different values. When users interrogate an individual comment, they are prompted to rate comments for how much they agree with and respect it. The size of the comment's dot on the map then grows when people with different values than the speaker respect and/or agree with it, facilitating users in seeking out comments that resonate widely.²

SYSTEM DESCRIPTION

In Washington state, measures can be added to the ballot by the legislature or, with sufficient signatures, can be submitted directly by citizens. Some of the nine 2010 measures were hotly contested, particularly Initiative 1098 which would have instituted Washington's first-ever state income tax, taxing income over \$200k (single) or \$400k (joint), reducing property taxes by 20%, and increasing tax credits for businesses. Other measures were opaquely worded and confusing, such as two

²MetaViz [4] is also closely aligned in spirit, though not design strategy. It seeks to trigger critical thinking in political blogs by computationally identifying and exposing the metaphors upon which discussants are drawing.

irreconcilable measures for privatizing the sale of liquor. Unfortunately, there were few places for citizens to actively work through the various arguments and claims being made by campaigns and pundits, or hear the considerations of "everyday" people. This provided an opportunity to facilitate reflective public thought. Six months prior to the election, we partnered with Seattle City Club, a nonpartisan civic organization, with the goal of supporting a community-written voters' guide on the ballot measures.

ConsiderIt is the result of the design process undertaken to realize the *Living Voters Guide*. ConsiderIt was designed and developed by the first author, with assistance from the second author and a graphic designer. The second author also developed usage scenarios and conducted an early paper prototyping user study. The rest of the project team participated in decisions about features and helped test the system. ConsiderIt is implemented in Ruby on Rails, with extensive use of the jQuery Javascript library for client-side interactions. It is available under the AGPL open source license.

While ConsiderIt is applicable beyond electoral deliberation, we ground our system description in the LVG. We describe it in two parts: (1) supporting personal deliberation by crafting a position; and (2) exploring aggregated positions to understand the considerations that resonated most with other users. In each section, we use two personas to introduce the functionality, followed by a description of the components not covered in the scenarios. The personas' actions are composites based on the behavioral traces of real users from the LVG deployment who were considering the income tax measure I-1098. The points the personas include and author are actual user-contributed points. We dramatize the personas' motives based on two of the aforementioned communication styles in order to illustrate how users with diverse motives might appropriate the LVG. Specifically, Jim is an undecided voter who wishes to carefully weigh his options, while Maria is an advocate working on behalf of her cause, engaging in behaviors associated with communitarianism.

ConsiderIt: Deliberation through position crafting

Last week, a couple of Jim's friends argued over I-1098. It wasn't clear to Jim where he stood. This morning, he came across a link to the LVG while loading the home page at the Seattle Public Library and decides to see if it can help him sort out his thoughts. When Jim arrives at the LVG, he clicks the icon for 1098. Jim reads the official summary description of 1098 at the top of the page, but does not follow the link to the full description because he doesn't want to get bogged down in legalistic language. Under the description is a slider asking Jim's initial opinion, from strongly support to strongly oppose. He plays with the slider for a few moments, but ends up leaving it at "Neutral". Next he comes across a chalkboard slate with "pro" and "con" columns. In the margins, he sees four pros and four cons written by others. One of the Con points is similar to one his buddy mentioned: "*The state legislature may expand the income tax to the middle class in two years.*" Jim thinks this is important, so he clicks the arrow on the Con point to include it into his own list. In the margin, another Con replaces the one he included. He cycles through several more sets of Con points, including



Figure 1. Jim's position on 1098. The "chalkboard" in the middle contains the user's pro/con list. On the sides, four pro and four con points written by others are shown. Clicking on the white arrow includes the respective point into the user's list. Not shown here is a neutral description of the issue being considered—it is above the pro/con list.

two more. He then browses through the Pro points, including one. Then he notices a Pro point stating "Why do people say voting isn't necessary to expand the tax? It says, the leg. must vote and the people must approve. We have a say." This contradicts the Con he included earlier. He has trouble reconciling these, although the full description of the Pro cites the relevant passage from the measure in the text, which makes him more dubious about his friend's argument. He removes his friend's point from his list. After reading through a dozen more points, Jim notices that one of the most important points that his other friend brought up is not represented. He decides to add it and clicks "Add pro": "1098 would lower property taxes for struggling middle class families." He includes a hyperlink to a *Times* editorial in the description of his point in case anyone is interested in reading more. As he scrolls down, he notices a second slider prompting him to update his stance if it has changed. Looking at his pro/con list (Figure 1), he thinks "am I still neutral?", then nudges the slider about a quarter way towards "Oppose" because he's still skeptical. He likes being able to convey his uncertainty, rather than simply saying yes or no, even if he ultimately has to vote. He clicks "Finish".

Maria volunteers for a campaign called "Yes on 1098." She has been spending part of her afternoon scouring the internet for discussions about I-1098 in order to make the case for it. She comes across the LVG and wants to make sure that the "Yes on 1098" perspective is being forcefully represented on the site. Maria reads that the LVG is a guide to the 2011 Washington state ballot initiatives written "by people like you". She knows from experience that sites like this often contain poor quality information or present a lopsided perspective. She wants to make sure that the arguments from the "Yes on 1098" campaign are properly and prominently repre-

sented, so she goes directly to the 1098 page. Maria already knows 1098 inside and out, so she breezes past the description and grabs the stance slider and moves it towards 'Support' as far as it will go. When she scrolls down to the Pro/Con list, she starts by cycling through the existing pro points, including many of them in her list. She then skims the Con points to see what her opponents are saying about 1098. She sees a couple that aren't completely out of line, but most are bogus. After some consideration she decides to include one Con point: "Local companies will be forced to pay more state taxes than national companies like Wal-Mart." If there's ONE thing about this bill that could be slightly tweaked, that's it. She sees one Con written by a guy named Lawrence that claims that 1098 will cause wealthy people to leave the state. "That's not right", so she writes a pro point to counter it: "Bill Gates Sr. is one of the chief proponents of 1098, so despite what some people here are saying there's obviously some support for 1098 among the wealthy!" Thus inspired, she goes on to create several more Pro points, ending each one with "Vote Yes on 1098!"

Additional details

The primary components of the position crafting page are two stance sliders and a pro/con list. After the description, users are asked to take a stance signaling their initial support for the issue. The pro/con list follows. Users can contribute their own pros and cons, providing a summary (≤ 140 characters) and an optional long description (≤ 500 characters). If a long description is provided, a "read more" link is displayed for the point, which expands on click to show the full description. Each pro or con displayed in the margin has a white arrow pointing toward the user's pro/con list. Clicking on the arrow includes the respective point in the user's list, and the included point is replaced by another point that has not already been displayed to the user. Every point a user authors while writing his or her personal pro/con list is made available for other users to include. A second slider after the pro/con list prompts users to reflect on whether their stances have changed over the course of creating their pro/con lists. The sliders are linked so that moving one also moves the other. Though having people state their stances before the pro/con exercise may cause opinion anchoring, we chose to include the two sliders for the simple reason that in early user tests, participants found moving the slider and seeing the "support" and "oppose" text grow and shrink to be very engaging and was thus a good UX hook. At the bottom of the page is a "Finish" button that submits the user's position and redirects them to the results browsing page.

ConsiderIt: Exploring the guide to public thought

Thus far we have described how ConsiderIt facilitates personal deliberation augmented by the considerations others are making. In this section, we describe the system components that helps users explore and make sense of emerging public thought arising from these personal deliberations.

Returning to our personas, after **Maria** submits her position, she starts browsing the guide (Figure 2). She clicks on the bar representing other strong 1098 supporters to make sure that all the basic points are there. They are. Then she clicks on the strong opposers bar to see what arguments the other side favors. She is surprised to see them admitting that 1098

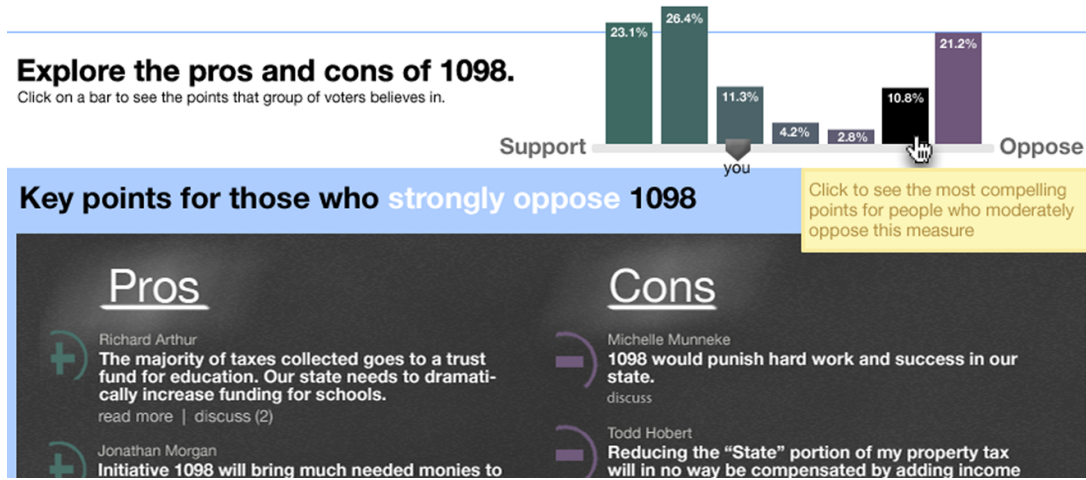


Figure 2. Browsing the guide. The current breakdown of support for a measure is shown in a bar graph to the upper right. The graph is interactive: clicking a bar changes the points shown in the pro/con list below the graph, revealing a ranked list of the most important points for those who took that particular stance. The initial list of pros and cons shown are the overall most salient. Not shown is a discussion forum below.

would provide needed resources for social services, but otherwise rolls her eyes at the points they find persuasive. Regardless, maybe the social services consideration will provide a good starting point for dinner conversation when she goes to visit her conservative in-laws next weekend. She clicks a button that automatically leaves a post on Twitter telling her friends and followers to check out LVG page for 1098, noting that none of the Cons were convincing.

Jim submits his position and is taken to a page where he sees a bar graph of the support and opposition to 1098. After reviewing the top-ranked points, he clicks on each of the seven bars to make sure he has given supporters and opposers the chance to put their best arguments forward.

Additional details

The results page presents an evolving guide to the salient pro and con points. First, the stance histogram enables people to see the distribution of support. The seven bars are discrete stance groups (e.g., weak support), derived from the continuous slider values that users selected to represent their stances. Second, the results page has a ranked list of popular points, again shown on a chalkboard slate of pros and cons. The ranking metric is described later. The most innovative functionality is that the stance histogram is interactive; clicking on a bar updates the slate of pros and cons to show the most important points for the users who took that specific stance. The results page also contains a threaded discussion forum. Each thread can be pegged to specific pro or con point by clicking a “discuss” link for each point shown in the pro/con slate.

We developed PointRank to amplify the points that resonate widely while still allowing for new considerations to emerge. The ranking determines the ordering of the pro and con points shown on the margin of the pro/con list during position crafting and to identify the most important points to display on the results page. There are three terms in PointRank.

Persuasiveness: how well this point convinces people to include it. It is the ratio of users who included it to the number

of users who viewed it.³ Because new points always start with the maximum persuasiveness score (the author includes it and is the only viewer), this metric helps new points gain visibility and mitigates the preferential attachment (rich-get-richer) problem [22] in asynchronous voting systems. Moreover, Bailey & Horvitz found that users of an enterprise ideation system wanted the ranking to incorporate the ratio between views and votes [3].

Diverse appeal: measures the degree to which this point appeals to both supporters and opponents, in order to highlight points that might surface common ground, rather than divisive points. For each of our seven stance groups (from strong support to neutral to strong oppose), we calculate the persuasiveness of that point for users who took that stance. The diverse appeal measure gives the highest score to points where there is equal persuasiveness across all stance groups, i.e., given an inclusion of the point, there is equal likelihood that the includer has taken any of the seven possible stances. Entropy directly measures this quantity: the points with the most diverse appeal maximize entropy, while the most divisive points minimize it. To formalize the calculation, let I_s be the number of inclusions of the point by users taking stance s , and let U_s be the number of unique users who were served the point and took stance s . Let $\alpha_s = \frac{I_s}{U_s}$ be the persuasiveness of the point for stance group s . Then the probability of the inclusion of the point by a user taking stance s , assuming that there are n stance groups, is given by $p_s = \frac{\alpha_s}{\sum_{i=1}^n \alpha_i}$. Entropy is then $e = \sum_{i=1}^n -p_i \cdot \log_n p_i$.

Raw appeal: the number of people who included the point. The raw appeal metric helps account for the confidence we have in the ranking. We are far more confident in a five-star rating for a restaurant based on a thousand reviews rather than on only three.

In our implementation, each of the three terms is normalized by calculating the respective point’s percentile rank against

³We logged all points a registered user was served while position crafting.

the other pros or cons for the issue. The final score combines the three terms linearly with equal weighting, an arbitrary choice made because it is not immediately obvious what the “optimal” weights would be. Future experimentation should help refine this choice. To facilitate the drill-down functionality into the considerations of e.g. strong supporters, the diverse appeal term is dropped, and the popularity and persuasiveness terms are based on the actions of strong supporters. In the LVG, scores were updated hourly.

PointRank was also designed to help protect against covert strategic manipulation of the results. However, we believe that the LVG deployment did not sustain high enough traffic to attract the attention of campaigns that might execute such an attack. We therefore relegate our untested explanation of PointRank’s resilience to a footnote.⁴

DEPLOYMENT DETAILS AND DATA

In the remainder of this paper, we present results from the LVG deployment of ConsiderIt. The LVG was launched on 9/21/2010 to a crowd of 150 at a Seattle City Club event. City Club seeded each ballot measure with one pro and con and led the outreach effort. Their nonpartisan reputation helped spread the word through liberal and conservative outlets statewide. We secured articles in the *Seattle Times* (9/27), *KIRO News* (10/5), the *UW Daily* (10/20), and the *Yakima Herald* (10/27), which drove most LVG traffic. Some websites linked to us from their homepage during the election season (e.g., the Seattle Public Libraries, public radio outlet KOUW). Furthermore, the LVG was used in several college and high school classes. Team members also reached out through email lists and social networking sites.

As an open deployment, we know little about the political attitudes of our self-selected population of users. On one hand, the *Seattle Times* drove the most traffic, and its comment sections are hyper-partisan. But perhaps only open-minded people followed the link. We cannot say, and we did not ask visitors to fill out a profile (to prevent the experience from feeling clinical). We can, however, provide some evidence in the form of metrics of use. Between 9/21 and 11/2, Google Analytics data shows LVG received 12,979 visits from 8,813 unique visitors. Ignoring the 6,082 sessions in which users visited only the homepage, users stayed an average of 10 minutes 39 seconds and visited 6.1 pages. Users from 134 Washington cities accessed the LVG (50.4% from Seattle). Traffic spiked on days of media coverage, but generally grew over time until election day (Figure 3). Many visitors opted to

⁴What if a group opposing a measure instructs its members to visit a ConsiderIt application to distort the results? They have at least two options: (1) they could add a number of cons masquerading as pros and instruct their members to include them in their lists, thus crowding out legitimate pros; and/or (2) they could add a number of inane or offensive pros and instruct their members to include them in their lists, thus making the opposition look bad. However, PointRank mitigates these attacks. First, the diverse appeal metric makes it hard to increase a point’s ranking without people who both support and oppose the measure including the point. One response is to instruct members to take a variety of stances, but that would act against their interest in showing overwhelming opposition to the measure. Second, the ability to click on a bar in the stance distribution enables people to see that the strong oppose group were the ones masquerading points or making inane or offensive points.

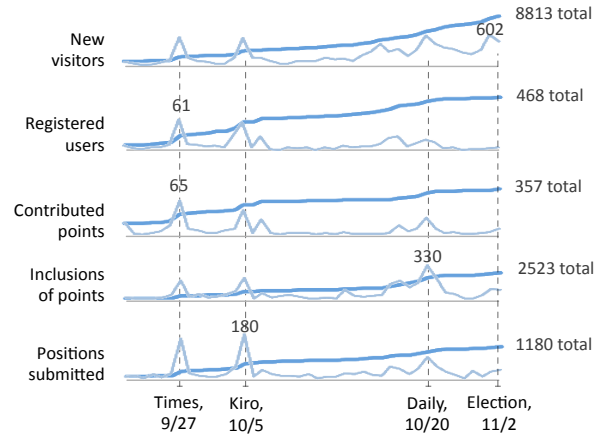


Figure 3. High-level trends. In every plot, each line operates at different scales: the dark line gives a cumulative count and the light line is the daily count. Highest daily peek and final counts are labeled.

simply browse the guide, where they could see the pros and cons that others had already submitted.

Our results draw on the following data sources.

Activity traces. The timestamped activity traces captured in the database are our primary data source (Figure 3). A total of 468 people registered and submitted a position on at least one measure. 184 con and 160 pro points were written by 147 users. The maximum points contributed by one user was 10. 298 users included these points 2,687 times into 678 unique pro/con lists (503 pro/con lists were left empty). The number of points written per ballot measure was skewed ($\mu = 38.2$, median = 29, [11, 113]). All actions taken by any member of the project team are excluded from the dataset.

User study. On the last days leading up to the election, seven Seattle-area residents were recruited from Craigslist and offered a \$20 Amazon giftcard for their participation in a lab study. Participants ranged in age from 18 to 47 ($\mu = 34.3$), all had some college experience, five were female, and they were racially diverse (two hispanic/latino, one asian pacific, two white, one black, and one multiracial). None had heard about the LVG. All but one had opinions about the ballot measures. The goal was (1) to get a sense of the perceived value of the LVG before the election had passed; (2) to learn how well users understood and valued the basic interaction mechanisms; and (3) to understand how participants reacted to the points others had submitted. Participants were not briefed about the nature of the site or our involvement in its development, and were free to interact with the site for a period of 35-40 minutes while thinking aloud. Though a researcher observed their activities and asked questions about their motivations, they were only prompted if they consistently bypassed some functionality. Audio and on-screen behavior were recorded. The first and second authors coded the data to identify aspects of the system that demonstrated ConsiderIt’s limitations and affordances, or that revealed interesting tensions arising from the design. In our results, we refer to lab study participants as P1 through P7.

Survey. During registration, we asked users for permission to administer a short post-election survey. 21.6% of the 250

willing users completed the survey. The average age was 42.5 ($\sigma = 17.1$), 50.3% male, overwhelmingly white, and high educational attainment (25% with postgraduate degrees). The questions were phrased as likert scale responses about their LVG experience. There may be a response bias: when we examined survey respondents' stances on two of the most ideologically polarized issues (1098 and 1053), we found that respondents exclusively took liberal stances (100%, 20/20), while those who did not take the survey took mostly liberal stances (60%, 156/260), while the general election returns were 36% liberal stances on the two measures.⁵ One dimension of bias in our population is thus clear: compared to the actual election returns, LVG users trended liberal; but our survey respondents were even more liberal than the LVG user population.

RESULTS

In this section, we synthesize data from our three sources to investigate how people appropriated the LVG. We start with our most important results, which quantify the extent to which users considered tradeoffs and engaged other users' points. This is followed by an analysis of survey data regarding the relative helpfulness of encountering points written by others. We then turn to data that helps us understand whether users actually shifted their opinions during their LVG experience. Finally, we draw on user study data to help provide an understanding about how users may have interpreted the considerations other voter segments were making.

Did the LVG encourage balanced consideration?

The power of pro/con lists is that they nudge the list creator to make sure they have thought about both sides of a decision. Despite this affordance, one might expect that few people participating in an open site during a contentious election would actually craft positions that acknowledge tradeoffs, given the partisan nature of our political discourse. However, **of the 678 positions that included at least one point, 41.4% of them included both a pro and a con.**⁶ Discounting the 148 positions with only one point, the frequency increases to 53.0%. Moreover, in the 40 cases where users authored more than one point for a measure, 45.0% of them wrote at least one pro and con. Thus, nearly half of those who were motivated to write multiple points decided to write both supporting and opposing points. Without comparative data, we cannot say whether the pro/con list format caused people to consider both pros and cons more than they otherwise would have. We suspect that because an unbalanced list is visually salient, the nudge to explore and potentially include points that challenge one's stance is particularly strong when a user is looking at an unbalanced pro/con list, with the points that others have written beckoning in the margin. Future work should explore this in a more controlled setting.

Balanced consideration is not just about considering tradeoffs, but also about whose points are being considered. In

⁵A glitch in our survey caused us to lose the link to the actual user in the database for roughly half of survey participants, so this data only refers to those for whom we retained the link.

⁶We restrict our analysis to positions where users included at least one point (57.4%) because despite our efforts at making the inclusion functionality prominent, it was clear from our user study that some users had difficulty discovering it.

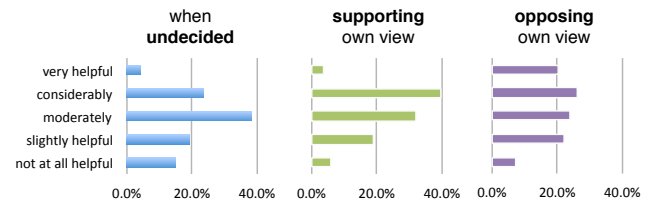


Figure 4. Helpfulness of reading points under different conditions, by fraction of survey respondents. Using pairwise 2-tailed Z-tests with $\alpha < .05$, we can reject the null hypothesis that the mean helpfulness of points is the same for those who are undecided vs. viewing opposing points, but not for those who are undecided vs. viewing supporting points ($p < 0.10$). There is no significant difference between the mean helpfulness of viewing supporting or opposing points.

ConsiderIt, a pro or con point may have been written by someone with a very different opinion on the issue. One might expect that few people would include points that were authored by people with whom they disagree. This is not the same as including both pros and cons: a given political stance can recognize both pros and cons for a given issue. For example, the recent bailout of the auto industry in the U.S. has distinct pros and cons for conservatives (preventing national security risk vs. state intervention) and liberals (saving local jobs vs. rescuing irresponsible corporations). In the LVG, **of the 599 positions where a user with a non-neutral stance included a point written by another user with a non-neutral stance, 33.7% included a point written by someone who took an opposing stance.** We suspect that the prevalence of users including points written by opposers has something to do with the pro/con list's structural nudge toward balance, as with including pros and cons. But another possibility is that ConsiderIt does not enable users to gain insight into a point author's political affiliation or stance on the issue beyond what they could infer from what the user wrote. In other words, *ConsiderIt does not provide group cues to activate political identity.* Social identity theory [39] would predict that had these cues been included, there may have been less diverse engagement because of internalized tendency to conform to a role (e.g., would someone with my political identity recognize a point written by the enemy?).

When was it useful to read others' points?

When we look at the inclusion metrics, we find that the stronger the stance that someone took, the less likely that person was to include both pros and cons (linear regression, $r^2 = .249$, $p < 10^{-12}$). Similarly, the stronger a stance, the less likely they are to include a point written by someone taking an opposite stance ($r^2 = .082$, $p < 10^{-5}$). But this does not imply that it was not helpful to read opposing points, or that it was useful to read supporting points. And it does not tell us whether undecideds actually found the points useful in establishing a position.

We turn to our survey to provide insight. We asked: (1) "For issues you entered undecided on, how helpful was it to read the points others wrote?" (2) "If you were leaning in one direction, how helpful was it to read opposing points?" (3) was the same as (2), except for supporting points. The results are shown in Figure 4. Viewing points was on average moderately helpful when undecided ($\mu = 2.83$, $\sigma = 1.08$). However, it was not as helpful on average relative to the help-

fulness of viewing supporting or opposing points when already having formed some opinion. One interpretation is that those who were undecided may have been more likely to be stymied by not knowing what information to trust, whereas those who already had an opinion had a stronger basis to navigate amongst the subjective voices. Seeing opposing points was seen as the most helpful ($\mu = 3.30$, $\sigma = 1.23$), possibly because they serve as a check on one's own decision. As P5 stated, *"I would probably read the opposite opinions to see if they would help me clarify my own opinion. It's more helpful to me to read people who disagree with what I'm thinking."* The expressed utility of seeing opposing viewpoints is somewhat surprising given that the literature on selective exposure suggests that many people are challenge averse. Perhaps our survey respondents were a biased selection of diversity-seekers (to use Munson's term [35]). Interestingly, while there was no significant difference in the mean helpfulness of seeing supporting or opposing points, there was much greater variance for opposing points. This may reflect diversity in the normative preference for encountering difference (also found in a different context by [35]).

Did LVG participants shift their opinion on issues?

Despite users' proclivity to acknowledge tradeoffs, recognize points by opposition, and find reading points by others helpful, it still does not necessarily follow that the experience impacts users' opinion on the issue. This is difficult to examine, but a few sources of data together suggest that there was in fact an impact on opinion:

Stance slider changes. The stance sliders were instrumented to log any time the user manipulated them (starting midway through the deployment). Looking at the change between the last slider manipulation before and after users started including points into their pro/con list gives an indicator of the effect of pro/con list-building on opinion. In 36.7% of the positions, users only manipulated a slider before crafting their list, while 25.2% only did it afterwards. The remaining 38.3% manipulated a slider before and after creating a list. Of these, 47.4% shifted their opinion by least one stance group (e.g., from "strong" to "moderate" support), with 56.7% strengthening their stance, 30.0% moderating, and 13.3% flipping sides and strengthening (e.g., from "moderate support" to "strong oppose").

Position updates. A few users updated a position on multiple days ($n=25$). Of these, 36.7% strengthened their existing stances on the issue, 54.5% flipped sides, and 9.0% moderated their stances. 53.3% of these users also included at least one new point into their positions during the update, suggesting that they were not just updating their stances to reflect some external impact, such as talking to someone outside of the LVG who influenced their opinion.

Self-reported opinion changes. 46.3% of survey respondents claimed that they actually changed their stances on at least one measure while using the LVG, with 56.0% of these users saying that they switched from support to oppose or vice versa, 32.0% saying that they moderated their stances, and 12.0% saying that they strengthened their stances.

Although not unproblematic, this data together suggests that a sizable fraction of LVG users either strengthened or mod-

erated their positions during or between their deliberations. Some even flipped their support.

What do participants look for in the considerations of others?

The system's most innovative functionality for exploring public thought is the ability to drill down into the most salient pros and cons for different stance groups by clicking on the bars of the histogram. This enables users to gain insight into the aggregated sentiment of people with whom they have different views, or to reflect on the sentiments of people who have taken a similar stance. As P4 remarked, *"It's interesting to be able to see a voting segment and [see what they value] and watch the points that slip off their mental map."* With this functionality, we hope to show people that there are other people with stances different than their own who are nevertheless "being reasonable" by explicitly recognizing the legitimacy of both pros and cons.

In an unfortunate oversight, we did not log every time that a user clicked a bar in the LVG, so we cannot report on user behavior in the deployment and must rely on user study data to gain insight into how this functionality may have been used. P3 provides an illustrative vignette. New to the state, P3 was unfamiliar with the measures. During the study, he used the LVG to form an opinion on one of the liquor sale measures. He found that the con points were actually prompting him the most to support the measure because he disagreed with what appeared to be the dominant con argument. After crafting his position, he started exploring the bar graph. *"If I was nervous about which side I'm taking, I can go to the strongly opposing, and they do give some cons there, so if I didn't see a con I liked earlier on the [position crafting page], these are the ones that people [who took a strong opposing stance] actually use"*. Thus, he could give each voting segment the chance to put their best case forward, supported by the drill-down functionality. While doing this, he noticed that one of the pros he found most convincing was listed highly in the salient points for those who strongly opposed the measure. This caused him to reconsider his stance: *"If the person who is... all the way on the oppose, still has the same opinion as you do on the pros, it makes it a whole lot stronger of a foundation for thought because even people who are fully against it are still agreeing on that pro with you. So that does make it a lot more of a valid point they're making [about opposing]."* This suggests that ConsiderIt might help people find common ground. P5 shows the converse, while exploring the bars: *"[there are] pros and cons, until you get all the way down to strong oppose where there aren't any pros at all, which is pretty unusual because even the people who are strongly supporting it have a few cons... I would probably be more swayed by uh, people who are a bit more open... I just find it a little bit strange that there's not a single pro on that."* Our data, however, is too sparse to claim that this affordance for evaluating how seriously to take different voter segments is likely to be discovered and exploited in the wild.

DISCUSSION

Here, we step back and examine some of the problematic aspects of ConsiderIt and what we believe are promising next steps. We first discuss the tension between trust and identity when users engage points written by strangers without an

organizational affiliation. We then discuss aspects of point redundancy, including further techniques for encouraging users to factor other peoples' thoughts into their own opinions, and then scaling ConsiderIt-style deliberations through clustering and collaborative synthesis.

Whom and what do I trust?

Although statements of personal principle were common, preliminary analysis indicates that at least half of the points submitted to the LVG contained some kind of potentially verifiable statement of fact, such as references to numeric data, external information sources, the textual content of the measure, or specific predictions about potential outcomes of adoption or rejection. The feedback we received during the user study and in the survey all suggest that the biggest challenge users faced was evaluating the trustworthiness of these claims. This is a recurrent issue with nearly all "crowd" powered approaches [1, 26]. Making sense of claims seriously constrained the ultimate utility of the LVG as an *information resource*. We believe that an important future direction is to enable users to link information sources explicitly to individual points to allow fact checking. The platform providers can thus avoid accusation of strong-handed moderation [49] by not making a value judgment about the validity of the source—anyone could decide for him- or herself whether a link to a pundit's blog is a reliable source or not.

Tightly coupled with trust is identity. Almost immediately after raising the issue of trust, user study participants would comment that they wanted to know more about the point author. This is consistent with social translucence, which suggests that providing people with more information about other online users can help them make judgements about who and what to trust and how to engage [11]. Recall that it was not possible to learn more about point authors beyond their name. The obvious implication is to allow an author to flesh out a profile, and enable other users to explore the stances that a point author took in the LVG. But this would undermine our design decision to omit such information based on the hypothesis that the lack of information about political affiliation helps nudge people to consider the points of diverse others. The design implications we draw from social translucence [11] and social identity theory [39] thus appear to be in conflict here. Future research might examine the differences between versions of ConsiderIt that are identical except in their respective degrees of identity salience, to tease out the consequences of the tension between listening to diverse others and evaluating claims. This is important, as some researchers are creating interfaces that highlight ideological differences rather than downplay them [17]. We recommend that future ConsiderIt deployments enable people to fill out profile information so that users can get a sense of personality and life perspective, but steer clear of calling specific attention to political affiliations.

Self-expression, redundancy, and scale

Given that we are asking the "crowd" to generate pro/con considerations, one might expect a proliferation of duplicate points that make it difficult to identify unique considerations. While we have yet to complete an in-depth content analysis to identify the prevalence and nature of redundancy, we can

see from the relationship between point additions and inclusions over time that **users self-regulated: 50% of all points were contributed by day 15, whereas it took until day 30 to reach 50% of all point inclusions** (Figure 3).

But we have found redundancy to be a more complex phenomena than it appears. First, "duplicate" points can add value by elaborating on an existing argument, reframing it in less (or more) inflammatory language to appeal to different audiences, or revising it for clarity. Second, there is a tension between expressiveness and redundancy: in a study of opinion influence in online discussion, Price et al. found that being in the presence of arguments only indirectly influenced post-discussion opinion by influencing the arguments that the respective participant chose to express themselves [39].

We draw an implication from this tension: if we provide more opportunities for users to put what other people are saying into their own words, participants may be more likely to incorporate the thoughts of others into their own opinions. We see two relevant extensions to ConsiderIt. First, when users include a point in their lists, ConsiderIt might prompt them to rephrase the point to their liking, starting from the original text. Second, taking a position might include the option of writing a summative position statement that weaves together the pros and cons to explain the user's stance.

While the ability to support self-expression through rephrasing might help reduce redundancy, further techniques are necessary if we are to scale ConsiderIt-style deliberation beyond what we attracted in the LVG. We see two broad steps in moving forward. The first step is to cluster points thematically via crowdsourcing or NLP, building a network of points linked by the semantic distance between them. This would allow us to better serve a diversity of points during position crafting, as well as enabling a range of recommendation opportunities, such as recommending a con point about the decision's impact on education after the user includes a pro point about education. Creating these similarity graphs is one step toward *synthetic deliberation*. In order to communicate at scale, we believe we will ultimately need to craft approaches in which users band together to express a larger point. For example, with clustered points, we could ask authors who have made very similar arguments to try to collaboratively merge their points. Or we could ask users to summarize clusters of points. Regardless, the design space around these synthetic mechanisms is vast and challenging; even wikis are confusing to many users. By first developing the core deliberative mechanisms of ConsiderIt, we feel we are in a better position now to extend them in potentially confusing but high value directions.

CONCLUSION

Our primary contribution is a novel technique that extends Franklin's pro/con technique for personal deliberation into the realm of public deliberation. ConsiderIt augments personal deliberation by enabling users to adopt points others have contributed, to contribute new points in turn, and to use the results of these personal deliberations to expose salient points. People can publicly deliberate and reflect on the thoughts of others without emphasizing direct discussion. Though we developed ConsiderIt for a regionally-focused

election, ConsiderIt may be applicable for a wider range of organizations seeking to engage loosely organized publics of citizens, employees, and customers.

Our secondary contribution is to demonstrate ConsiderIt's utility for deliberation amongst the general public. We take the stance that a strong aggregate indicator of utility is a system's discretionary use in a real deployment. Users spent more than ten minutes on average at the LVG, many of them returning multiple times. Our claim about utility is simply that the nature of usage suggests our approach finds a sweet spot in a challenging design space, given that discretionary use by the general public has been an elusive target for deliberative systems. Beyond demonstrating discretionary use, we showed that LVG users often engaged in normatively desirable activities such as crafting positions that included both pros and cons and recognizing points written by people who do not agree with them. Still, there were some clear challenges for users, like the ability to decide what to trust.

Finally, we have highlighted opportunities for controlled experimentation to better understand the design decisions, such as whether the pro/con list structure itself actually nudges people to consider tradeoffs more than a simple list of considerations. More broadly, we designed ConsiderIt to provide extra support for deliberative communication, but tried to accommodate other communication styles, such as the communitarian and liberal individualist. Future research might help more precisely characterize the appeal of ConsiderIt to those adopting different communicative styles.

ACKNOWLEDGEMENTS

Thanks to our partners at Seattle City Club, Diane Douglas and Jessica Jones; Andrew Hamada and Michael Toomim for help with design; Sam Kaufman and Alena Benson for help with the user study; and Scott Bateman, Eric Baumer, Michael Bernstein, Amy Bruckman, Scott Golder, Kurt Luther, David McDonald, and three anonymous reviewers for feedback on the paper. This work is supported by NSF grant IIS 0966929.

REFERENCES

1. B. Adler et al. Assigning trust to wikipedia content. In *WikiSym*, 2008.
2. C. Anderson. Deliberative, agonistic, and algorithmic audiences. *Int. J. of Communication*, 5, 2011.
3. B. Bailey et al. A case study of a grassroots innovation pipeline within a large software company. In *CHI*, 2010.
4. E. Baumer et al. Fostering critical and creative thinking about metaphor in political blogs. In *CHI*, 2010.
5. Bell, W. and Labaree, L (eds). *Mr. Franklin: A Selection from His Personal Letters*. Yale University Press, 1954.
6. I. Beschastnikh et al. Wikipedian Self-Governance in Action. In *ICWSM*, 2008.
7. J. Conklin et al. 15 years on from gibis. In *HYPERTEXT*, 2001.
8. L. Dahlberg. Democracy via cyberspace. *New Media & Society*, 3(2), 2001.
9. N. Diakopoulos et al. Videolyzer: quality analysis of online informational video for bloggers and journalists. In *CHI*, 2009.
10. J. Dryzek. *Liberals, Critics, Contestations*. Oxford Press, 2002.
11. T. Erickson and W. A. Kellogg. Social translucence. *TOCHI*, 7(1), 2000.
12. S. Faridani et al. Opinion space: a scalable tool for browsing online comments. *CHI*, 2010.
13. J. Fishkin. Strategies of Public Consultation. *Integrated Assessment*, 6(2), 2006.
14. A. Forte et al. Increasing Decentralization in Wikipedia Governance. In *HICSS*, 2008.
15. N. Fraser. A Contribution to the Critique of Actually Existing Democracy. *Social Text*, (25/26), 1990.
16. D. Freelon. Analyzing online political discussion using three models of democratic communication. *New Media & Society*, 12(7), 2010.
17. M. Gamon et al. Using Blogs to Provide Context for News Articles. In *ICWSM*, 2008.
18. R. Garrett. Echo chambers online. *J. of CMC*, 14(2), 2009.
19. J. Gastil. *Political Communication and Deliberation*. Sage, 2008.
20. J. Habermas. *Structural Transformation of the Public Sphere*. MIT Press, 1991.
21. S. Hansen et al. Wikipedia and the possibility of rational discourse. *The Information Society*, 25(1), 2009.
22. B. Huberman. *The Laws of the Web*. MIT Press, 2001.
23. S. Isenmann et al. IBIS: a convincing concept . . . but a lousy instrument? In *DIS*, 1997.
24. S. Iyengar et al. Deliberative public opinion in presidential primaries. In *Voice and Citizenship Conference*, 2004.
25. N. Jankowski et al. *The Prospects for Political Renewal Through the Internet*, chapter Internet-based political discourse. 2002.
26. A. Kittur et al. Crowdsourcing user studies with Mechanical Turk. In *CHI*, 2008.
27. S. Knobloch-Westerwick et al. Looking the other way. *Comm. Research*, 36(3), 2009.
28. T. Kobayashi et al. Selective exposure in political web browsing. *Information, Communication & Society*, 12(6), 2009.
29. K. Kraemer et al. Computer-based systems for cooperative work and group decision making. *ACM Comput. Surv.*, 20, 1988.
30. T. Kriplean et al. Supporting Active Listening and Grounding on the Web through Restatement. In *CSCW Horizons*, 2011.
31. P. Lange. Commenting on comments. In *Society for Applied Anthropology*, 2007.
32. K. Luther et al. An online collaboration environment for citizen scientists. In *CHI*, 2009.
33. M. MacKuen et al. Resolute partisanship or reflective deliberation. *Am. J. of Pol. Sci.*, 54(2), 2010.
34. P. Muhlberger. The virtual agora project. *J. of Public Deliberation*, 1(1), 2005.
35. S. Munson et al. Presenting diverse political opinions. In *CHI*, 2010.
36. D. Mutz. *Hearing the Other Side*. Cambridge, 2006.
37. S. Park et al. Newscube: Delivering multiple aspects of news to mitigate media bias. In *CHI*, 2009.
38. W. Parker. *Teaching Democracy*. Columbia, 2003.
39. V. Price et al. Normative and informational influences in online political discussions. *Comm. Theory*, 1(16), 2006.
40. D. Schuler. *Online Deliberation: Design, Research, & Practice*, chapter Online civic deliberation with E-liberate. 2009.
41. B. Shanks et al. *Online Deliberation: Design, Research, & Practice*, chapter A module for parliamentary procedure. 2009.
42. F. Shipman et al. Formality considered harmful. *JCSCW*, 8(4), 1999.
43. S. Shum. *Visualizing Argumentation*. Springer, 2003.
44. S. Shum. Cohere: Towards web 2.0 argumentation. In *COMMA*, 2008.
45. A. Sukumaran et al. Normative influences on thoughtful online participation. In *CHI*, 2011.
46. C. Sunstein. *Republic.com 2.0*. Princeton Univ. Press, 2009.
47. M. Trice. *The Ethics of Emerging Media*, chapter How We Comment on Web Journalism. 2011.
48. K. Wahl-Jorgensen. Letters to the editor as a forum for public deliberation. *Crit. Stud. in Media Comm.*, 18(3), 2001.
49. S. Wright. Government-run Fora: Moderation, Censorship and the Shadow of Control. *Brit. J. of Politics and Int. Relations*, 8, 2006.