©Copyright 2011
Travis Kriplean

Encouraging Reflective Discussion on the Web

Travis Kriplean

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

University of Washington

2011

Program Authorized to Offer Degree: Computer Science & Engineering

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Travis Kriplean

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

_____

Alan Borning

Reading Committee:

_____

Alan Borning

_____

Andrew J. Ko

_____

David W. McDonald

Date: _____

In presenting this dissertation in partial fulfillment of the requirements for the doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to Proquest Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, 1-800-521-0600, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature_____

Date_____

University of Washington

Abstract

Encouraging Reflective Discussion on the Web

Travis Kriplean

Chair of the Supervisory Committee:
Professor Alan Borning
Computer Science & Engineering

Trust in our democracy has been eroding, undercutting our ability to confront difficult problems. Meanwhile, our public discourse is fragmented and polarized. In this dissertation, I argue that these issues are inextricable and that one promising approach for transcending them is to create communication technologies that help publics find common ground upon which they can organize and take action. To inquire further, I have created two new interactive technologies, ConsiderIt and Reflect, that help facilitate large public discussions on the web by encouraging people to consider tradeoffs and listen to each other. ConsiderIt guides people to reflect on the tradeoffs of an issue through the creation of a personal pro/con list, with the twist that they can adopt points others have already contributed. ConsiderIt then surfaces the most salient pros and cons overall, while also enabling users to drill down into the key points for those who support, oppose or are undecided. Reflect, on the other hand, modifies online comment boards by creating a space next to every comment where people can succinctly summarize and restate points they hear the commenter making. Commenters thus know whether they are being heard and understood, and readers can demonstrate understanding. Analyses of multiple field deployments of ConsiderIt and Reflect illustrate that participants in the deployments used these technologies in normatively desirable ways, suggesting that the design direction may have wide traction amongst the public. I also contribute a tiered theoretical framework for designing effective public deliberation that puts these contributions into context. Whether these and similar technologies can be combined to create more effective organizational structures that build trust and counter fragmentation over the long term remains to be seen.

TABLE OF CONTENTS

Page

List of Figures iii

Chapter 1: Introduction 1
1.1 Theory and design for public deliberation 6
1.2 Statement of contribution 22

Chapter 2: Supporting Reflective Public Thought with ConsiderIt 25
2.1 Introduction 26
2.2 System description 28
2.3 Deployment details and data 35
2.4 Results 37
2.5 Discussion 43
2.6 Living Voters Guide, 2011 46
2.7 Conclusion 47

Chapter 3: Promoting Listening on the Web with Reflect 49
3.1 Introduction 50
3.2 Interfaces for Grounding and Listening 51
3.3 Reflect: Listening Through Restatement 55
3.4 Field Deployments 60
3.5 Discussion and conclusion 81

Chapter 4: Lessons Learned for Designing for Reflection 83
4.1 Introduction 84
4.2 Lesson 1: Articulating utterances as length-limited, semantically constrained Points provides leverage for lowering barriers to reflecting, involving more people, and surfacing common ground 85
4.3 Lesson 2: column-based layouts can structurally make reflective activities understandable, actionable and enforceable 99

Chapter 5: Conclusion and Future Work 104
5.1 Future work: toward robust collective intelligence 105
5.2 Intellectual contribution 111

Appendix A: Interview data 113
A.1 Based on your experience writing Reflect bullet points... 114
A.2 Based on your experience having a comment you wrote Reflected by someone else... 116
A.3 Based on your experience reading a Reflect-enabled discussion... 117
A.4 Based on your experience participating in and possibly leading community discussions (not necessarily with Reflect)... 119

Bibliography 122

LIST OF FIGURES

Figure Number — Page

1.1 The core ConsiderIt interaction. 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. 4

1.2 An excerpt from a Reflect-enabled Slashdot story. The entire comment section is embedded to the right in the margin. Readers of the digital version can zoom in to peruse the whole discussion. Those who printed the pdf will need to fetch their magnifying glasses. 5

1.3 Differences between the self-Actualizing Citizen and the Dutiful Citizen which our civic institutions assume. Taken directly from Bennett [9]. 7

1.4 Three models of democratic communication and their associated behavioral indicators. Taken directly from Freelon [37]. 12

1.5 Affordances for different communicative orientations: indicates whether the interface discourages (purple) or encourages (green) the respective behavior when diverse communicators are present. Light saturation indicates implicit support, while dark saturation indicates explicit support. The boxes describe briefly some aspects of the interface that explain its categorization. This is not absolute: anyone is able to adopt any of these behaviors across the different interfaces. 15

1.6 Evident frustration with the incentive to participate. 20

1.7 PeerToPatent (http://peertopatent.org/), which crowd-sources finding prior art on submitted patents for the overburdened USPTO, is a great example of acknowledging participant contributions in the agency's process. Shown here is a prominently displayed widget on its frontpage. 21

2.1 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. . . 29

2.2 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. 32

2.3 *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. 36

2.4 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. 40

3.1 Mechanics of the Reflect interface. 55

3.2 The first use of Reflect in the wild, on the Living Voters Guide. In this exchange, Vin first restates a point he thought James was making, and then responds to James' comment. James received an system-generated email indicating that someone restated a point he made. He followed the e-mail's link and clarified his point to correct Vin's misinterpretation. The exchange illustrates a major issue with asynchronous listening: Vin replied to James' comment before James came back to verify and correct Vin's restatement of his point. On one hand, anyone browsing the forum after James responded to Vin's summary would benefit by avoiding Vin's misinterpretation. But Vin's own reply now makes less sense. 57

3.3 An excerpt from a Reflect-enabled Slashdot story. The entire comment section is embedded to the right in the margin. Readers of the digital version can zoom in to peruse the whole discussion. Those who printed the pdf will need to fetch their magnifying glasses. 62

3.4 The discussion prompt for the Amazon Mechanical Turk discussion. LIke the Slashdot figure, a major portion of the Reflect-enabled discussion is embedded on the right-hand side of the screenshot. Zoom in on the digital version to read it. 70

3.5 An except from one of the discussion pages for the Wikimedia Strategic Planning Initiative. As before, the full discussion is embedded on the right-hand side of the screenshot, and can be read by zooming in on the digital version. 74

3.6 Screenshot from a CCF discussion where JS is restating points of participants in a discussion. Unfortunately, there is not a longer screenshot available due to a database problem resulting in the loss of a number of blog posts on the CCF page. . 77

4.1 An early prototype of Reflect that used a single wiki as the comment summary. . . 94

4.2 PopVox (http://www.popvox.com), the most similar system to ConsiderIt. Side note: observe the "appreciate" framing of the "like" button, a much cleaner articulation of listening functionality than the "like" button (which overloads judgment and listening), although the need to click the heart icon detracts from this conceptual distinction. 95

4.3 An example graph-based layout of a discursive space, from one of the most popular examples of this technique, http://debategraph.org. 100

4.4 2d visual presentation of comments, from http://www.nytimes.com/interactive/2011/05/03/us/20110503-osama-response.html. . 101

5.1 One possible task in a summarization workflow. Summarizer is asked to incorporate Reflect restatements into a partially created discussion summary. 106

ACKNOWLEDGMENTS

There have been many people who deserve to be acknowledged for supporting me in the past six years. Of course, my parents Dan and Linda, and my partner Eva Ringstrom who has had to put up with many strange lines of inquiry, long days and nights of work, not to mention frequent requests to give extremely early feedback on nearly all of designs. And there are numerous friends and colleagues at the UW and elsewhere that have made graduate school possible and, moreover, extremely enjoyable.

I have had the pleasure of working with academic collaborators like Michael Toomim, Deen Freelon, Jonathan Morgan, Lance Bennett, David McDonald, Ivan Beschastnikh, Andrew Ko, Mark Zachry, Evan Welbourne and Paul Waddell on various projects. While much of the research we collaborated on is not explicitly included in this dissertation, it has greatly shaped what I have focused on and how I look at the world. For example, I was attracted to study Wikipedia because I was intrigued by its deliberative norms and means of enforcing them, which I had the freedom and support to study with David and Ivan [10, 63, 62]; the attention paid to observing vast amounts of Wikipedia deliberations greatly influenced the design inspirations underlying Reflect and ConsiderIt, even if the connection is not direct. Classes taught by Andrew Ko (design thinking) and Lance Bennett (political communication) have also been seminal.

And of course, my advisor Alan Borning has been supportive at multiple levels throughout the past six years. He has created the space for me to figure out and explore what I'm interested in. Sometimes, this entailed allowing me to do work that was not explicitly funded, but which I sensed was valuable. I am grateful for this patience. This is to me the most important and significant outcome of my PhD experience: learning what I'm passionate about and finding people who share those passions.

In the last couple years, I have met and worked with some very engaging people outside of the university. These include Jessica Jones and Diane Douglas of Seattle City Club, Andrew Hamada

of Reinspire.Me, Rob Hohn of IdeaScale, Bill Schrier of Seattle City Government, Eugene Kim of Groupaya, John Spady of King County Community Forums, and Rob Malda, co-founder of Slashdot. I would like to also thank Cliff Lampe who connected me with Rob Malda at Slashdot, which led to the much needed Reflect deployment. And Tim Vroom, an engineer at Slashdot, helped make the trial a reality.

I would also like to thank the faculty and students who participated in the CHI 2010 and CSCW 2010 doctoral colloquiums, as well as the NSF sponsored 2011 SoCS doctoral consortium. These experiences were instrumental in helping me sharpen the intellectual scope of this work.

Finally, I'd like recognize some of the National Science Foundation grants that have supported me throughout my PhD years: IIS-0966929, IIS-0705898, and IIS-0811210.

Chapter 1

INTRODUCTION

Democratic governments need to have some degree of trust established with the citizens it purports to represent in order to function effectively as a decision-making body and arbiter of the peace. Unfortunately, in America, there has been a severe drop in trust of its government over the last several decades, accompanied by a general sense that our institutions are no longer working for us [70]. As of late 2011, trust in Congress is at an all time low of 9%.[1] In a 2010 Rock the Vote Poll, 62% of those polled said that they would not vote because "no matter who wins, corporate interests will still have too much power and prevent real change".[2] It seems that the only thing that we can agree on is that our government no longer represents its people. This state of civic destitution is problematic if we are to address the challenges confronting our cities, states, country and world.

Can we rebuild trust and reform the institutions that tie us together? In order to do so, I argue we need to foster greater deliberation amongst the citizenry. This is not an uncommon perspective, particularly in academia. It is routinely theorized that engaging others in *public deliberation* has many benefits, such as building trust, including diverse others in public decisions, detecting ill-conceived ideas, and making high quality decisions [33, 44, 41].

Unfortunately, wide, effective public deliberation has been elusive. 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 [77]. The use of communication media reflect these tendencies, where people are highly attracted to content that aligns with their prior beliefs, even those who do not actively avoid challenging content [40, 56, 57, 76]. 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 [80]. 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.

Discussion on the web reflects these high-level issues with political communication. In this

---

[1]Trust in congress varies according to the poll, but has recently been between 9-14%. The 9% figure comes from "Congressional approval at all-time low of 9%, according to new CBS News/New York Times poll": `http://www.cbsnews.com/8301-503544_162-20125482-503544/`

[2]This came from a talk Lawrence Lessig did for his book [70]. The source poll is at `http://www.rockthevote.com/assets/publications/research/2010/2010-rock-the-vote-nationwide-baseline.pdf`

dissertation, I address two problems that I argue are central for adding wide deliberative capacity to our public discourse. First, web discussions are often fraught with alienating personal attacks and flaming [3], crowding out those who might be more open to productive exchange across social boundaries. Second, discussions often grow to include so many different people, each wanting to voice their own opinions, that it becomes difficult to identify actionable high level takeaways. Thus, while social media may offer new infrastructure for people everywhere to connect, to share ideas, to come together and create wonders like Wikipedia (thousands of volunteers work together to document reality itself, using a consensus-based deliberative model!), issues of civility and scalability are curtailing the efficacy of social media to derive value from public discourse. They are curtailing our ability to take action together and relate to each other in a more productive fashion. Even in Wikipedia, project-wide deliberation is straining the limits of civility and scalability. Civility and scalability may appear to be separate issues, but I argue in this dissertation that a single design strategy provides leverage on addressing both at once.

> **Design strategy:** Create web interfaces that (1) nudge people to reflect on the issue and/or other participants' thoughts and (2) re-present evidence of this reflection in order to help participants (a) find greater satisfaction in participating in or reading the discussion and (b) come to a better understanding of the higher level issue with which participants are grappling.

The core of my dissertation is a pair of novel systems that embody this strategy:

1. *ConsiderIt* focuses people on thinking through the tradeoffs of a proposed action, such as a ballot measure in an election, by inviting participants to craft a pro/con list (Figure 1.1). The innovation is that the pro and con points that others have already authored are shown in the margins, which people can include into their own lists. The pro/con considerations written by others thus becomes available for reflection by other participants as they also consider the tradeoffs of the action. Through this process, ConsiderIt can then surface the most salient pro and con points, helping participants understand the points that resonate most strongly with other participants. ConsiderIt also enables users to drill down into the salient points for different segments of the population ("What were those strong opposers actually thinking?").

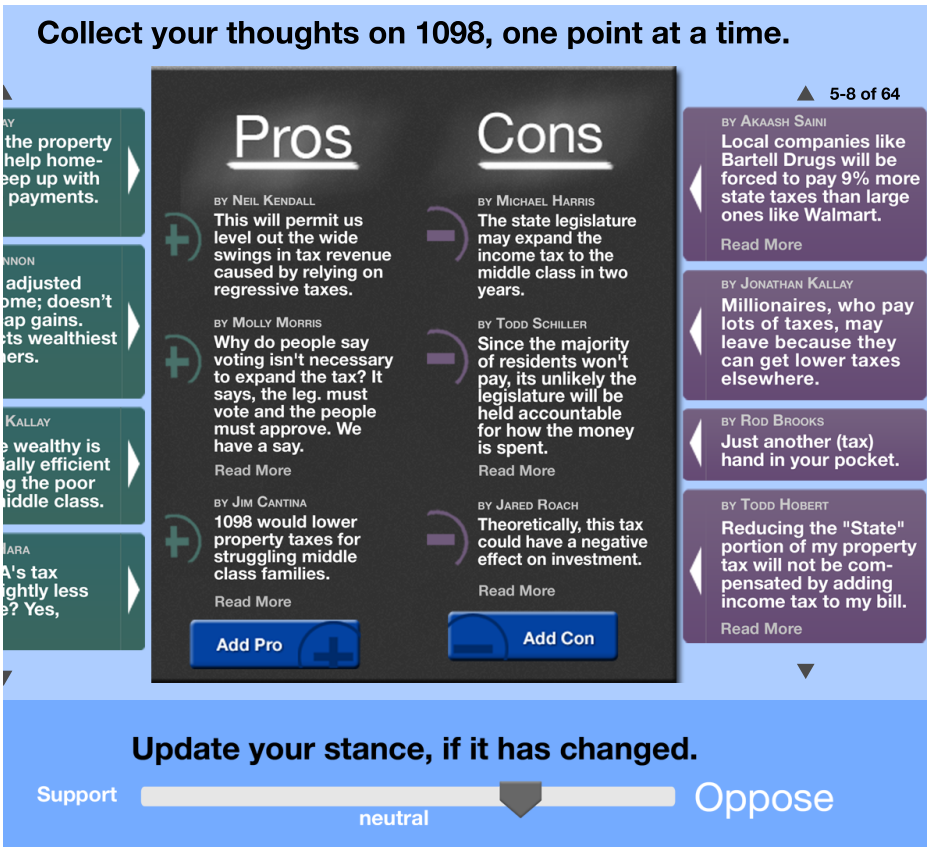


Figure 1.1: The core ConsiderIt interaction. The "chalkboard" in the middle contains the user's pro/con list. On the sides, four pro and four con points written by others are shown. Clicking on the white arrow includes the respective point into the user's list. Not shown here is a neutral description of the issue being considered–it is above the pro/con list.

People can thus gain insight into the considerations of people with a different perspective than their own, rather than making assumptions based on caricatures of them, an experience that contrasts favorably with the zero-sum shouting between representatives of the extremes that are often portrayed in mass media as "debate".

2. *Reflect* nudges participants in online forums to listen to each other, to seek mutual understanding about what commenters are trying to say (Figure 1.2). The goal is to mitigate the prevalence of knee-jerk responses and increase the likelihood of finding common ground. Re-

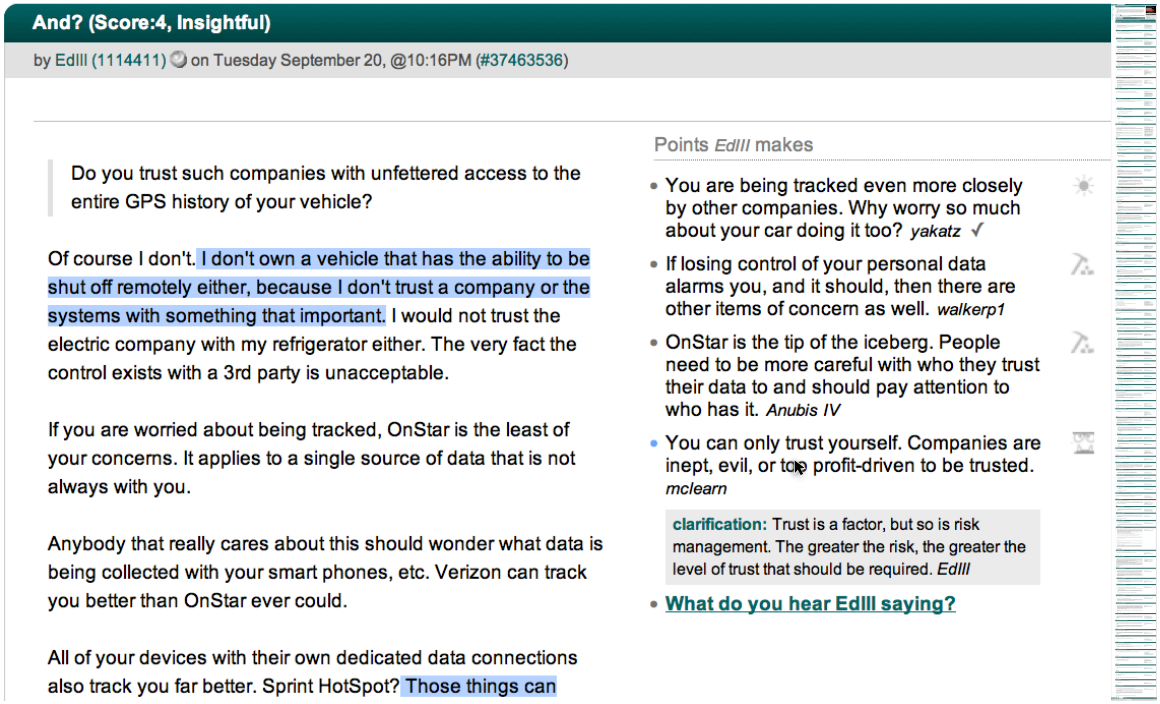


Figure 1.2: An excerpt from a Reflect-enabled Slashdot story. The entire comment section is embedded to the right in the margin. Readers of the digital version can zoom in to peruse the whole discussion. Those who printed the pdf will need to fetch their magnifying glasses.

flect introduces a space to the right of every comment where readers are invited to restate a point that they hear the commenter making. Any reader can do this, and each restatement is represented in a bulleted list. Each of these bullet points represents an act of listening. Reflect also enables commenters to confirm whether a summary of their comment is accurate. If necessary, commenters can also clarify their point. Future readers are then exposed to the original comment, interpretations of that comment, and clarifications of those interpretations, creating a richer commenting environment. Aside from creating a more empathetic commenting environment, the output of the reflective mechanism, restatements, also provide an intermediate step toward summarizing the full discussion, critical for deriving long term actionable outcomes from large discussions.

I have evaluated ConsiderIt and Reflect through multiple field deployments. ConsiderIt is the

result of a design process initiated with the high-level goal of supporting a community-written voters guide for Washington State deployed in the contentious 2010 U.S. election (and less contentious 2011 election). Likewise, Reflect has been deployed in a number of real world situations, including the popular technology news discussion site Slashdot and a participatory governance effort by Wikimedia. This emphasis on field deployments reflects my 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. Demonstrating utility for wide usage has been a major hurdle for prior work in the civic space, where academics tend to create systems that implement their own societal utopia, often requiring a level of investment in arcane activities that are far from engaging the motivations of most people. By emphasizing field deployments, I am able to ensure that the systems I design are at least more likely to have wide traction, as well as gaining insight into the utility and affordances of a social system operating in a contested and often vitriolic domain.

1.1 Theory and design for public deliberation

The core of this dissertation is one chapter each about Reflect and ConsiderIt, covering the system design and rationale as well as results from multiple deployments. Before diving into these chapters, I will go into more detail about the intellectual frame that I have been developing in tandem with my design inquiries. In the process of developing this frame, I will cover prior work and explicitly state my original intellectual contributions.

1.1.1 The emerging networked public sphere

Any organized group confronts decision-making challenges when setting policy and choosing how to allocate scarce resources, such as how to balance a budget, who should replace the outgoing board president, or what policies should be implemented to incentivise public transportation. Although many organizations still prefer one-way, bounded communication channels, there is a surge of excitement and high level leadership in governing organizations to use social media to assemble large, loosely connected publics to help deliberate, whether these constituents are citizens, cus-

Actualizing Citizen (AC)	Dutiful Citizen (DC)
Diminished sense of government obligation—higher sense of individual purpose	Obligation to participate in government centered activities
Voting is less meaningful than other, more personally defined acts such as consumerism, community volunteering, or transnational activism	Voting is the core democratic act
Mistrust of media and politicians is reinforced by negative mass media environment	Becomes informed about issues and government by following mass media
Favors loose networks of community action—often established or sustained through friendships and peer relations and thin social ties maintained by interactive information technologies	Joins civil society organizations and/or expresses interests through parties that typically employ one-way conventional communication to mobilize supporters

Figure 1.3: Differences between the self-Actualizing Citizen and the Dutiful Citizen which our civic institutions assume. Taken directly from Bennett [9].

tomers, a student body or employees. Large businesses like Starbucks and Dell are creating portals for customers to suggest ideas and let other people vote on them; internally, organizations like Best Buy, Microsoft and the U.S. federal government are establishing internal processes for employees to suggest and execute ideas for improving how the organization functions [6]. Political parties are tentatively exploring crowdsourcing various input, from platforms to video on campaign events. And sites like Facebook are gathering feedback on terms of service changes in their Site Governance initiative. Beyond these *consultations*, more radical organizations have participatory processes at their core, such as in Wikipedia where contributors struggle to conduct distributed self-governance through massive deliberations [10, 63, 34, 45], or political movements that lack central organization.

At the same time that we are seeing these changes in organizational possibilities, people are demanding more interactive and collaborative relationships with the organizations with which they identify. This is particularly true for younger generations who increasingly prefer to share knowledge (rather than defer solely to experts), express their views through peer networks (rather than just in the isolation of the voting booth or the opinion poll), address issues through informal networks (rather than joining more structured interest groups), and participate in formulating action (rather than following the lead of top-down authorities). Bennett [9] describes this as a paradigm shift in citizenship itself, with newer generation actualizing citizens (ACs) seeking a place in civic culture

alongside more conventional dutiful citizens (DCs) born in an earlier social and communication era (See Figure 1.3).

These cultural, organizational and technological developments can be seen as the emergence of a *networked public sphere* [8]. The public sphere is the space in which people "speak to each other in their relationship as constituents about what their condition is and what they ought to do as a political unit" [8, 177-8], and may take place anywhere (e.g. coffeeshops or newspaper editorials). This is a reference to Habermas' [44] conceptualization of different realms of social life in modern societies, the other two spheres being the state sphere (the activities of the government), and the private sphere (familial relationships, labor, markets, exclusive civil associations). The high-level political function of the public sphere is to criticize and discipline state authority through the "vehicle" of public opinion [43]. But the means by which this function has been fulfilled, to greater or lesser success, has shifted dramatically through time, often catalyzed by disruptive communicative technologies. After the fall of feudal society, the "bourgeois" public sphere emerged through a rise in literacy and wide circulation of newspapers. As the press grew, and radio and television were invented, the "broadcast" public sphere took hold, characterized by mass mediated public opinion. In the past several decades, the steadily increasing commercialization of mass media threatens its independence from the private sphere and hence dilutes its ability to play the "watchdog".

The broadcast public sphere is now being augmented and sometimes supplanted by a networked public sphere, catalyzed by the affordances of internet technology to cheaply and easily publish media, alter published content, and hyperlink. Benkler usefully characterizes this shift in terms of the necessary functionality that the public sphere must play in order to serve its role as a check on state power in a democracy. Specifically, he sees the public sphere as requiring the means by which there can be (1) a wide "up-take" of issues of public concern, (2) mechanisms for filtering out non-salient concerns, and (3) the means by which to synthesize concerns such they can become "condensation points" for collective action. In the networked public sphere, the means by which these functionalities are fulfilled are distributed based on the attention dynamics of citizens acting on their own accord across the web, as opposed to being consolidated by the gate-keeping, mass media organizations of the broadcast era. An example of this functionality at work is what precipitated Trent Lott's resignation as Senate Majority Leader following racist comments. A relatively minor blog kept pushing the story after Lott's comments were ignored by the mass media (wide up-

take). The story was picked up by "A-list" blogs, initially liberal and then conservative (filtered for relevance), who collectively raised a ruckus (consolidation point for action). This drew the attention of the mass media and ultimately lead to Lott's resignation. Overall, this networked public sphere has great potential, but also significant opportunity for great harm, such as political balkanization and viral misinformation in the absence of trusted gate-keepers.

1.1.2 Deliberation in the networked public sphere

I argue that this networked public sphere has yet to develop strong deliberative capacity to support collective action in the absence of formal organizational hierarchy. This is constraining the networked public in finding common ground across political boundaries, as well as undercutting the wisdom of the actions for which a public may decide to rally around and advocate.

I see two general classes of applications where adding deliberative capacity would be greatly beneficial.

First, creating communication channels that bridge the state and public sphere in a way that bypasses mass media. The goal is to create a sort of *socket* between a networked public of Actualizing Citizens with the cautiously interactive approaches being experimented with by governmental agencies that are organizationally and culturally shaped to engage with Dutiful Citizens. For example, government is not equipped to effectively engage citizens interested in participating outside of the election cycle. Consider public hearings. Because public hearings are only elicited after a plan is well-formulated, citizens are put in a reactive position and led to employ polarizing discourse during their 2-3 minute speaking slots [99, 52, 15, 48]. There is generally no dialogue or question answering. Busy and thoughtful people become discouraged and do not return [52]. Government officials are subject to many short statements of variable quality without an obvious way to synthesize the comments, gauge support for statements, or any real motivation to consider it [48]. Often, citizens and officials see the whole process as a spectacle, a gesture to satisfy legal mandates [52, 48]. We can design communication spaces that help citizens articulate coherent positions together, and that integrates more cleanly into an evolving policy-making cycle. A second example of this type of application is enabling mass heterogenous protests to more effectively communicate an accurate portrayal of their activities and goals to a broadcast mass media unequipped to interpret what is

happening on the ground.

Second, and more radically, new forms of organization that eschews hierarchy and emphasizes deliberation and consensus building might be supported. Wikipedia is an example, as is the Occupy Wall Street movement. While these efforts have great energy behind them, identifying points of consensus, or articulating differences, amongst potentially thousands of participants can create barriers for these organizations in making decisions and taking action.

To conceptualize why deliberative capacity in the public sphere is not being afforded, and to highlight design opportunities, I have been developing a three-level intellectual framework.[3] The first two levels regard how a communication space is designed for a loosely networked population of individuals and organizations to communicate their ideas, opinions and action items with each other. The last level regards how the communication is embedded in the larger decision-making context:

1. Encourage constructive discussion by nudging participants to consider the tradeoffs inherent in the issue being discussed, reflect more deeply about what other participants are trying to say, and introspect about one's own values. The goal is to foster more inviting discussions that focus on building toward solutions rather than cutting each other down and alienating participants. A representative sensitivity at this level is *What affordances can be provided to help a discussion from devolving after a commenter makes a personal attack?*

2. Enable discussions to scale better and create actionable output around uncovered common ground. Turn scale into an advantage, rather than a liability. Empower participants to express themselves as a group, or aggregate input in a way that enables high-level patters to emerge. A representative sensitivity at this level is *How can wide public discussions be summarized accurately, especially in the presence of strategic communication by interest groups?*

3. Design the institutional context in which a discussion takes place such that the outcomes of a discussion can be integrated into decision-making processes. A representative sensitivity at

[3] I first articulated many of these ideas through a consideration of work in public administration, political science, and communications (see [61, 60] for fuller, though dated, explanations of the framework).

this level is *What types of commitments do decision-makers need to make regarding how input is accounted for in order to motivate members of the public to actually participate?*

This framework is an initial attempt to create a design-oriented conceptual framework for mass participatory governance. It tries to move beyond HCI and CSCW's more traditional focus on small-group decision-making embedded in hierarchical organizations [58] to target wider, more distributed community discussion. It is an initial sketch for framing a design space that can be improved upon, both conceptually and empirically.

1.1.3 A framework for designing for networked deliberation

I will now address the three levels of this framework. My dissertation focuses primarily on the first level, engages the second level, and does not substantially contribute to the third.

Encouraging constructive discussion through reflective communication

Forums on issues of public importance often appear to be the site of drive-by trolling and flaming, alienating people who may be inclined to civilly engage with others. It is rare to come across a civil discussion about a contentious issue amongst an ideologically diverse group, or at least that is a common perception by internet users [26]. Based on my personal observations of many discussions across the web, discussions frequently degenerate into knee jerk responses and personal attacks when people do not feel that they are being heard, understood or respected as legitimate participants. Moreover, there is a natural tendency to downplay arguments and evidence that does not conform to our prior beliefs. Oddly, evidence that challenges one's preexisting beliefs sometimes reinforces and strengthens the original belief [94]. We humans are naturally defensive and protective of our beliefs rather than being prone to evolve them given new knowledge.

Despite these challenges, I argue that it is vital to mitigate some of these natural tendencies and encourage more reflective engagement: considering tradeoffs inherent in an issue, seeking to understand others through empathetic listening, and being able to step back and introspect about one's own beliefs and source of those beliefs (metacognition). With these behaviors, we will be in a better position to think together because we will be able to better find opportunities to connect our ideas together and extend them, or, conversely, understand better exactly where our values diverge.

Model of democratic communication	Indicative metric
Liberal individualist	Monologue
	Personal revelation
	Personal showcase
	Flaming
Communitarian	Ideological fragmentation
	Mobilization
	Community language
	Intra-ideological questioning
	Intra-ideological reciprocity
Deliberative	Rational-critical argument
	Public issue focus
	Equality
	Discussion topic focus
	Inter-ideological questioning
	Inter-ideological reciprocity

Figure 1.4: Three models of democratic communication and their associated behavioral indicators. Taken directly from Freelon [37].

User interface interventions might improve the situation at a couple levels. First, interfaces can serve as an aid for learning more reflective behaviors, and reminding us to be more reflective when it is easiest to forget (such as when we get irritated). 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 [79], even though most people appear to desire only news stories that back their own beliefs [76]. Second, perhaps participants in web discussions are in fact often being attentive and reflecting, but our interfaces do not explicitly capture and reveal it; reflections may simply be invisible. We can design interfaces that make those reflections visible.

The reflective behaviors I seek to support are roughly aligned with the values of deliberative democracy where decisions are rationally justified and relevant stakeholders have their voices heard and incorporated [95]. However, deliberation is just one communicative ideal for democracy, and it sometimes comes into conflict with others. Social scientists have begun to draw on political theory to juxtapose and measure these conflicting ideals [37, 4, 25, 97]. Freelon [37] maps out

three such ideals and their associated behavioral indicators within online spaces (see Figure 1.4). Aside from the deliberative style, liberal individualism [29] is an egocentric perspective celebrating self-expression over listening, manifesting in monologues and even 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 daily newspapers, both of which offer broad freedom of expression but which are abrasive and low in responsiveness [68, 96]. Communitarianism [36, 94] 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 [40, 56, 57, 76]. And while some people do not go out of their way to avoid challenging material, they still must encounter it first, an activity which platforms like Facebook and Google Reader do not encourage.

While I respect the importance of these different communicative ideals, most, if not all, broadly used communication interfaces implicitly or explicitly support the non-deliberative styles, likely crowding out those who wish to deliberate and stunting the development of deliberative skills in those who do not yet possess those skills. Some scholars and practitioners have developed tools and methods 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 [33]. Most systems work in this domain has tried to translate aspects of these multi-day, resource intensive deliberative efforts into online tools [50, 51, 75]. They tend to centralize control over the agenda and delimit acceptable contributions, such as by instituting Robert's Rules of Order [85, 87]. 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 (e.g. Toulmin argumentation [74, 55, 17], QOC [73], IBIS [83, 14, 22, 49, 20, 21, 86]), and incorporate them into a graphical visualization of the issues, arguments, positions, and so forth (see [53, 14] 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 [49, 88, 90], often requiring trained facilitators to be successful [23].[4] 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.

To add deliberative capacity to the networked public sphere, I argue that not only must deliberative styles of interaction be supported in more casually used interfaces, but that they must be supported in a way that also accommodates aspects of the other styles that do not undermine deliberation. If we can attract the attention of diverse communicators, the interface can gently nudge people toward reflective activities, such as weighing tradeoffs 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 [72] has shown that the communicative styles that people adopt are context sensitive, and we know that interfaces can help shape context [93].

Reflect and ConsiderIt are both designed with mechanisms consistent with deliberative principles at the core, but which can also engage the motivations of those adopting different communicative orientations (see Figure 1.5). For example, In ConsiderIt, creating a pro/con list is a deliberative activity, as is including pros and cons written by others. But contributing pros and/or cons might originally derive from a communitarian motivation to make sure that one's own side is adequately represented in the communication space.

Thus far I have positioned this work with respect to a deliberative perspective on communication, while using the term "reflective" quite often. This is because I do not fully subscribe to the "deliberative" value set. In my reading of how scholars use the term deliberative, too often is it associated with a strong emphasis on process, consensus and cold, hard logic, whereby we analyze and calculate the optimal choice, excluding other ways of knowing, such as telling stories [81]. I believe that this weakens the concept's potential to inspire transformative designs. Our problems do not stem from imbalance between our rational mind and our emotional mind, where we just need to discipline ourselves to use our reason to struggle out from beneath our primitive emotions in order to reach enlightenment. It is not that those who agree with you are rational and those who do not

[4] Recent work such as Cohere [14], Pathfinder [71] and Videolyzer [26] show promise by relaxing some of the constraints of the argumentation scheme.

Interface		***Communicative behaviors***		
		Liberal individualist (monologue, personal revelation, flaming)	**Communitarian** (mobilization, in-group identification, out-group exclusion)	**Deliberative** (rational argument, reciprocal listening, emphasis on public good)
	Online forums (e.g. comment boards, blogs, idea platforms)	emphasis on speaking (comment), reverse-chron ordering, crossposting, linear list of postings	crossposting, +1 thumbs up, blogrolls, banning users	difficulty listening to comments by many, bias against positive feedback
	High investment deliberative systems (e.g. based on Roberts Rules of Order)	some story telling		turn-taking, argument mapping, facilitation
	Lightweight consideration (e.g. ConsiderIt)	adding pro/con points, taking stance	reading/including points recognized by people with similar stance, taking stance	reading/including pro/ cons of people with diverse perspective, can't see political affiliation

Figure 1.5: Affordances for different communicative orientations: indicates whether the interface discourages (purple) or encourages (green) the respective behavior when diverse communicators are present. Light saturation indicates implicit support, while dark saturation indicates explicit support. The boxes describe briefly some aspects of the interface that explain its categorization. This is not absolute: anyone is able to adopt any of these behaviors across the different interfaces.

are emotional and deluded, and our task is simply to insist on reason in order to win the other side over: we are all subject to the polarizing warp of ideology on what we believe to be true. Rather than being a hinderance, our emotional mind is actually a well trained and critical component of our intelligence, assimilating far more information from past experiences than our rational minds can juggle at one time, and sending us emotional indicators based on these past experiences when we enter a new situation or hear someone telling us something [69]. There is even evidence that our rational mind often trails our emotional brains, retroactively justifying the decisions that our emotional minds have already decided for us [69]. Intuition is not to be scoffed at and dismissed, even when it cannot be articulated. The only way to address and counteract the distorting influence of polarizing ideology is to try to become conscious of our thought patterns and emotional state.

This is why I insist on "reflection". Reflection connotes a more holistic looking ever inward, deeper, probing issues, oneself, and others to greater depths, open to rational and emotional inquiry. I see reflection as closer to Buddhist conceptions of mindfulness, than a cold process of analytical

dissection. While the reflective behaviors I seek to inspire may align with deliberative norms, I believe that the distinction between reflection and deliberation is important as a design strategy. With a deliberative orientation, I believe we will always be, unreasonably, stuck blaming those who we believe are being unreasonable, while the ship sinks. With reflection, we might be able to design ways that encourage people to transcend differences and work toward a better future by better understanding the roots of belief systems.

Handling scale: listening to whole discussions

With attention to reflective mechanisms, we may be able to improve the quality and tenor of discussions and inspire more people to participate. But even if individuals are listening more to each other and reflecting on their own beliefs, there is still a major problem that limits the ultimate utility of these discussions: the sheer scale of a discussion can easily prevent participants from being able to understand, build on, and endorse or argue with what others have already said. Moreover, the coordinators of the participatory governance effort can have difficulty sorting through and extracting themes from the dialogue, highlighting innovative suggestions, identifying possible points of consensus, and responding. Governments in particular often lack the resources necessary to fully engage constituents.

This can be seen as a fundamental problem in the division of communicative labor imposed by blogs, comment boards, letters, and public hearings: people speak individually, proliferating their comments, and then at the end, if the necessary resources are available, representatives from the organization summarize the main points raised in order to identify actionable input or points to which decision makers might respond. We need to reformulate this division of labor so that communicators themselves become responsible as a group for collaboratively distilling, linking together, relating and summarizing their thoughts. Only then will organizations be able to efficiently derive benefit from public discussion.

This is the crux of my design strategy: the output from people's use of reflective mechanisms can itself be used to help summarize, link, and relate the statements of many. Reflections, such as inclusions of pro and con points in ConsiderIt, or restatements of others' points in Reflect, can be aggregated and repurposed 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 on how often and by whom the points are adopted, while in Reflect, the comment-level restatements can be used to power discussion-level summaries. By designing systems that hinge upon participants' use of reflective mechanisms, we can at once foster more welcoming and high quality discussion, but also better identify actionable outcomes from the discussion.

> **Design strategy:** Create web interfaces that (1) nudge people to reflect on the issue and/or other participants' thoughts and (2) re-present evidence of this reflection in order to help participants (a) find greater satisfaction in participating in or reading the discussion and (b) come to a better understanding of the higher level issue with which participants are grappling.

Essentially, my design strategy is to crowdsource aspects of professional facilitation into reflective mechanisms. Complex protocols and trained facilitators are our current response to managing large discussions and making sure they are constructive. Facilitators are trained to make sure that people are being heard and understood, for reframing sticky points, and for creating summaries. In other words, facilitators employ reflective techniques - summarizing, reframing, distilling, relating, mapping, clustering, composing, filtering - that provide evidence to speakers that they have been heard; this evidence is, at the same time, fed back into the conversation for use by other people to make sense of what is going on. But facilitators are rare and expensive. By crowdsourcing aspects of professional facilitation as core components of communication interfaces, we may better scale discussion and develop better communicators. The web is better suited to capture and repurpose distributed demonstrations of understanding made by many people than any other communication medium. By crowdsourcing facilitation, we can enable groups of people to better refine and evolve expressions of their beliefs in such a way that more people are able to participate and understand.

Related work. Outside academia, in commonly used web interfaces, only meager progress has been made toward reformulating the division of communicative labor. Progress has been limited to social filtering technologies. For example, idea platforms have emerged in the last couple years as arguably the most dominant civic communication platform. Consider Ideas For Seattle. Ideas For Seattle was launched by Mike McGinn during his bid for Mayor of Seattle. The site allows anyone to

submit an idea about "the future of our city" and allocate 10 votes to other ideas. Idea platforms are largely glorified threaded comment boards, with the notable addition that people can vote on ideas. The assumption is that the best ideas will bubble to the top with enough people participating.[5] By including people in the process of ranking ideas, these systems are starting to engage the theme of crowdsourced facilitation.

Another line of work takes the current division of labor as given and builds tools for post-processing discussions. One line of research creates "tools for rulemakers" to sort through comments, applying automated techniques for identifying near-duplicate comments (evidence of form letter derivation), and code them for interesting content [65, 64, 89]. Another line of research focuses on systems that enable communities to benefit from discussions by distilling the important aspects of them into "distillates" [1, 78] so that the takeaways of a large discussion can be found in the resulting "bramble". Both of these approaches are expert systems designed for use by professionals to handle the structural challenges of figuring out what many people were saying. Underlying each of these approaches is the stance that the ideal case would be automatic summarization (see e.g. [78, p. 139]). While fully automated summarization would be very useful, it would only address issues of scaling, not reflection. The design strategy I have laid out tries to address scale as a secondary outcome of encouraging more reflective discourse.

Few other systems align with the design strategy I have laid out here. OpinionSpace [32] is closest. It plots on a two-dimensional map the individual comments in a web forum, based on the commenters' responses to a short value-based questionnaire. From the start, therefore, it is nudging people to reflect on their beliefs. 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, helping participants make sense of what is being said. 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.

[5] www.ideasforseattle.org. The Obama administration has led a number of efforts employing similar systems (Citizen Briefing Book [7], Open for Questions [8], Dialogue on Open Government [1, 9]), and all federal agencies followed in order to implement his open government directive. They are seeing wide use outside of government, originating with Dell IdeaStorm and spreading from there. Participation varies, but idea platforms frequently get hundreds of ideas and opinions, and thousands of votes and comments.

While the designers of OpinionSpace have not articulated the design strategy I have described here, I see their system as another example of a technology in line with its principles.

Embedding discussions in meaningful contexts

The focus of my dissertation work is primarily on encouraging reflective discussion, and secondarily on creating the means by which we can scale discussions (the prior two sections). However, I see Reflect and ConsiderIt as building blocks for creating organizations that prosper based on their participatory governance efforts. With participant-summarized discussions, large discussions might actually have actionable output that can be incorporated into a higher-level decision-making processes. For example, a news site might incorporate points raised in a discussion into their decision about followup stories. Or a city government might host a series of discussions about a proposed policy, incorporating feedback at various key points in its lifecycle and showing how the feedback impacted the bill. In other words, to build anything of significance from Reflect, ConsiderIt, and other communicative technologies for supporting participatory governance, it is vitally important to embed the discussions in effective institutional relationships. Successful participatory governance efforts need to co-design the communication technology and the social context of its embedding in decision making processes [104]. In this section, I briefly outline two important aspects of institutional design for participatory governance.

Institutional commitments

A prevailing motivation for civic participation is to make your voice heard by the governing organization [92]. Unfortunately, most governance efforts are not designed from the beginning to listen responsibly to the collective input, never establishing what to do with the input before asking for it [92]. Worse, sometimes the whole thing is a farce, or the only "listening" that happens is what they wanted to hear in the first place - cherry-picking for politically expedient input [92]. Institutions must go beyond passively listening; they also need to actually demonstrate evidence that they are listening, otherwise participants will not recognize that they are listening. And they need to have a plan to do so. The framework suggests two forms of commitments to listening that governing organizations can make, each of which produces evidence of listening.



Figure 1.6: Evident frustration with the incentive to participate.

Committing to participate in the discussion. Governing organizations can commit staff to provide feedback *while the discussion is still ongoing*. And be very explicit about how much staff capacity they are dedicating. For example, staff might commit to responding to a certain number of statements per day (potentially based on the amount of activity surrounding the statements). Dedicated staff are the voice of the governing organization in the day to day operations of the socket. Their actions mean something categorically different than interactions with unaffiliated strangers. Reflect provides real leverage here: staff might be able to indicate listening by adding summary bullet points without needing to commit to an immediate action in the response.

Committing to integrate the input with the respective decision making process. Commit-

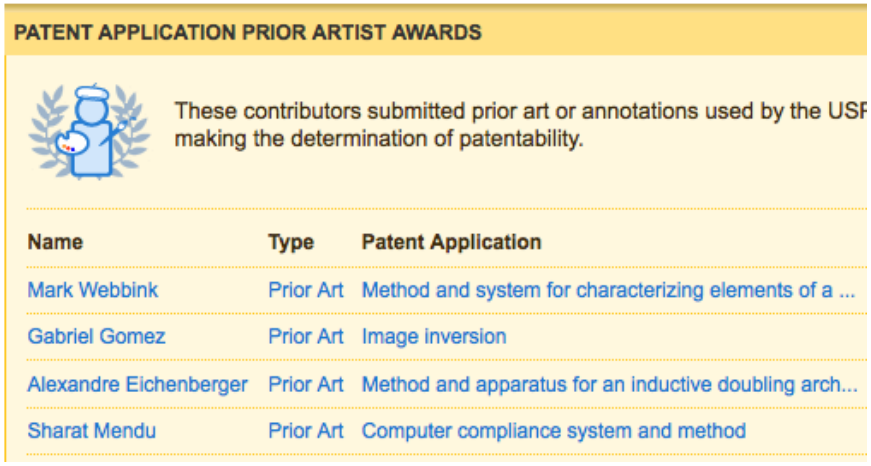


Figure 1.7: PeerToPatent (http://peertopatent.org/), which crowd-sources finding prior art on submitted patents for the overburdened USPTO, is a great example of acknowledging participant contributions in the agency's process. Shown here is a prominently displayed widget on its frontpage.

ments can be made regarding how the input will be incorporated into the larger decision making process. In preliminary discussions with City of Seattle officials, we learned that there are two key points in the decision making process where a socket would be particularly valuable: (1) early when the issue is first being explored, and (2) close to a final decision. Different outcome commitments can be secured: early in the process, there may be a commitment to set up a public meeting between city council and authors of influential statements; later in the process, influential statements may win the rights to be included as part of official voting pamphlets [41].

The commitments that organizations make should be explicit in the interface, and evidence of listening by organizations should be prominent. This will help participants have clearer expectations about possible interactions and outcomes so that they can frame their participation accordingly. Possibly the best interface component I have seen for demonstrating listening is PeerToPatent (Figure 1.7). But there are subtleties that depend on application. An example of this cropped up as part of our collaboration with IdeaScale, one of the leading idea platform vendors. In early 2010, IdeaScale was tapped to provide the software and hosting for all federal agencies to collect input on being more open (see http://opengovtracker.com/). One of the features that many idea platforms

had implemented, including IdeaScale, was a method of treating each idea like a ticket that could be marked as "implemented" or "in planning". Then, "your ideas completed" was listed prominently on the front page. Across the board, the federal agencies needed this feature removed. That listening mechanism implied too much commitment to action. The institutional listening mechanisms were too heavy-handed.

Reaching publics

If a governing organization is attempting to gather input from its publics, it needs to have an outreach effort in place. For example, the City of Seattle has a fairly well established protocol for reaching out to different constituencies using mixed methods, such as leveraging community centers, libraries, and mass media. Just as few will attend a public hearing held in an inconvenient location, it will be more difficult to attract various publics if the socket is digitally isolated. There needs to be a social media outreach strategy as well. A couple methods can be used. First, making it easy to participate via Facebook or Twitter allows people to speak behind their established identity. Moreover, it allows people to become advocates of the socket to their own networks. For example, when participants endorse a statement, they might choose to publish it to their Facebook feeds, thus alerting their friends and possibly pulling these friends into the portal as well. This can help engage diffuse publics. Second, relevant civic efforts and blogs can host widgets on their sites to advertise engagement opportunities to their constituencies. This will create many entry points to the dialogue. None of these considerations are terribly innovative but these design issues are vital to the success of participatory governance efforts.

1.2 Statement of contribution

I claim to make three intellectual contributions in my dissertation research:

1. A design frame that conveys a strategy for addressing challenges in participatory governance.

2. Two novel systems that demonstrate the aforementioned design strategy by centralizing reflective mechanisms: ConsiderIt and Reflect.

3. Field deployments of the two systems in real situations for discretionary use. Analysis of activity patterns, interviews and a user study suggest that participants frequently used the reflective mechanisms; moreover, the use was usually consistent with the designed intent.

Overall, these contributions are illustrative of what I call **generative social system research**. Research is focused more on innovation and field deployment, rather than empirical certitude, in order to create and inquire about new futures.

> **Design strategy:** Create web interfaces that (1) nudge people to reflect on the issue and/or other participants' thoughts and (2) re-present evidence of this reflection in order to help participants (a) find greater satisfaction in participating in or reading the discussion and (b) come to a better understanding of the higher level issue with which participants are grappling.

Using this design strategy as my thesis would overclaim what I can prove empirically. I have focused on field deployments, which do not lend themselves to experimental control. Thus the empirical evidence I collected through deployments are not appropriate for claiming that the reflective mechanisms actually enabled participants to find *greater* satisfaction or come to a *better* understanding of the higher level issues. Instead, I demonstrate the following thesis:

> Web interfaces that centralize the use of mechanisms encouraging participants to reflect on the issue and/or other participants' thoughts are engaging for general use and participants will frequently use them in normatively desirable ways.

Throughout the rest of this dissertation, I will focus on describing the two systems and demonstrating the above thesis with respect to each of them.

The core of the dissertation are two co-authored conference papers, which culminated work started in 2009. Those two chapters have been largely left intact, including use of "we" instead of "I", except for some additional empirical data and elimination of some material overly redundant with the introduction and conclusion of this dissertation.

This work was not completed independently. Beyond those recognized in the acknowledgments, I want to recognize a few collaborators who greatly impacted the intellectual contributions contained herein:

1. **Michael Toomim.** Michael has been my primary sounding board from high level framing of the problems, all the way down to design iterations on Reflect and ConsiderIt. Together we started reaching out to and conversing with IdeaScale and Seattle City Government, which ultimately resulted in the Engage group. He has been particularly important on the Reflect work. In addition to design support, Michael led the Mechanical Turk experiment analysis, which is written up in the Reflect chapter.

2. **Jonathan Morgan.** Jonathan has helped out in a number of ways on both Reflect and ConsiderIt. In fact, the name ConsiderIt was his idea, if I recall properly. Jonathan has been particularly important in undertaking qualitative analysis. He headed up the Living Voters Guide user study drawn on heavily in the ConsiderIt chapter. He also drafted the section on interview themes in the Reflect chapter, as well as drafting the user scenarios contained in the ConsiderIt chapter.

3. **Deen Freelon**. Deen's work on models of democratic communication has heavily influenced how I look at the affordances of communication interfaces. Beyond that, some of the framing in the ConsiderIt chapter derive from a paper that Deen and I coauthored. Deen also redrafted some parts of the introduction and related work section of the ConsiderIt chapter. Deen (and Lance) also created the survey used during the first Living Voters Guide deployment, which I draw on in the ConsiderIt chapter.

4. **Alan Borning & Lance Bennett**. Alan and Lance are the principle investigators for the NSF grant supporting our Engage group. Their high-level influence has been ever present, in particular Alan. But without Lance bringing together City Club and our group, there would be no Living Voters Guide. Both have contributed suggestions, and written several passages, in the chapters contained herein.

Chapter 2

SUPPORTING REFLECTIVE PUBLIC THOUGHT WITH CONSIDERIT

2.1 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 [35]

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 [33, 44, 41]. 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 [77]. 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 [40, 56, 57, 76]. 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 [80]. 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 people to drill down into the salient points for different segments of the population: *"What were those who strongly opposed this actually 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.

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 chapter, 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.

We thus present generative social system research: (1) we have *created a new interactive approach* for supporting public deliberation that has successfully weathered a fairly significant field deployment; (2) we describe a design strategy that *may spark new ideas* by other researchers; 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 presenting the design rationale for ConsiderIt in detail, we describe the LVG deployment, followed by an analysis of ConsiderIt's appropriation. We end with a discussion of challenges and future directions.

2.2 System description

In Washington state, measures can be added to the ballot by the legislature or, if sufficient signatures are collected, can be submitted directly by citizens. A number 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 $200,000 (single) or $400,000 (joint), reducing property taxes by 20%, and increasing tax credits for businesses. Other measures were opaquely worded and confusing, such as two 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 a design 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 is implemented in Ruby on Rails, with extensive use of the jQuery Javascript library for client-side interactions. As with all Rails applications, the code is organized in a Model-View-Controller architecture. ConsiderIt has been designed to be extensible to other domains beyond election deliberation. It is available under the AGPL open source license.[1]

While ConsiderIt is applicable beyond electoral deliberation, we ground our system description in the LVG. We describe ConsiderIt 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

[1]The Github repository is at `https://github.com/tkriplean/considerit`.

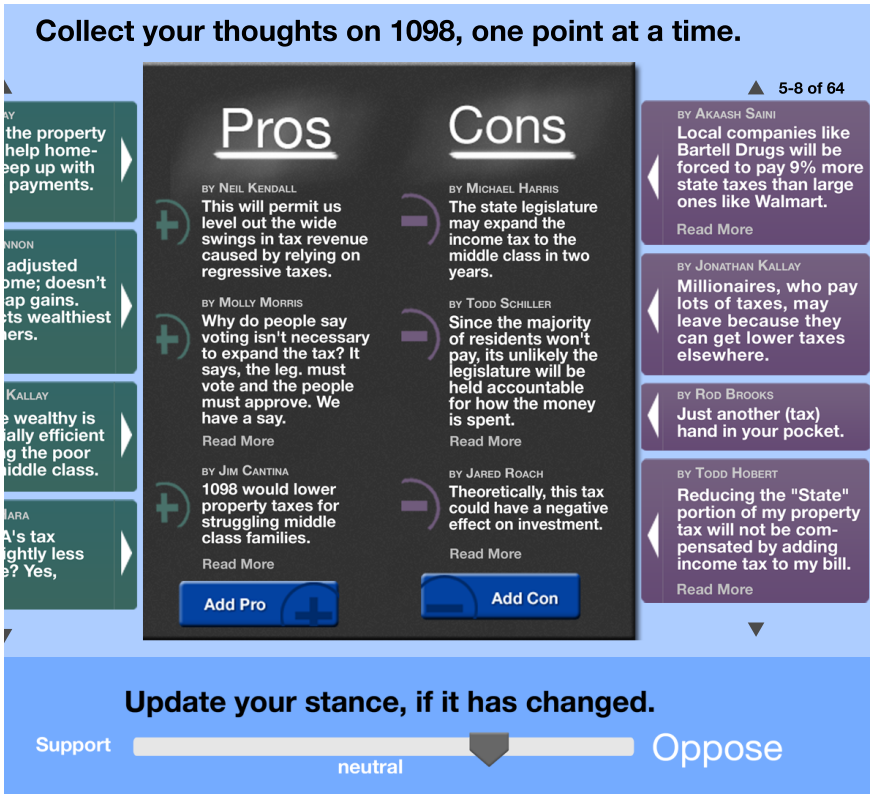


Figure 2.1: Jim's position on 1098. The "chalkboard" in the middle contains the user's pro/con list. On the sides, four pro and four con points written by others are shown. Clicking on the white arrow includes the respective point into the user's list. Not shown here is a neutral description of the issue being considered–it is above the pro/con list.

points. We dramatize the personas' motives based on two of the communication styles Freelon [37] describes 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.

2.2.1 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 (Figure 2.1). 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 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 2.1), 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 represented, 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

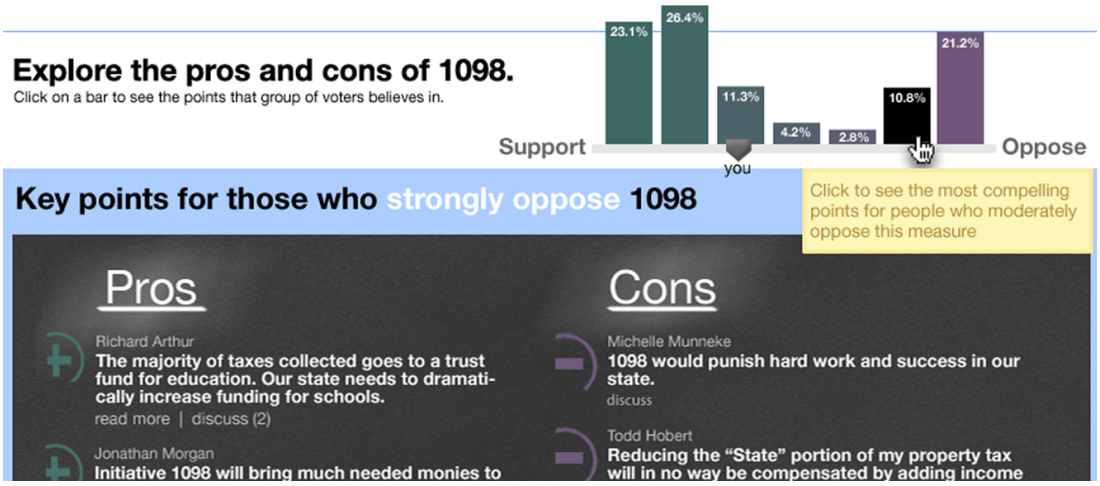


Figure 2.2: Browsing the guide. The current breakdown of support for a measure is shown in a bar graph to the upper right. The graph is interactive: clicking a bar changes the points shown in the pro/con list below the graph, revealing a ranked list of the most important points for those who took that particular stance. The initial list of pros and cons shown are the overall most salient. Not shown is a discussion forum below.

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.

2.2.2 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 2.2). 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 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 ever-evolving guide to the salient points on the issue. First, the stance histogram enables people to understand the distribution of support for the measure. 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 as a chalkboard slate of pros and cons. The ranking metric is described later in this section. 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. When clicked, if a discussion about the point has already been started, the window is scrolled to the associated thread; if not, a new thread is created for that point.

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.[2] 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 [47] in asynchronous voting systems. Moreover, Bailey & Horvitz found that users of an enterprise ideation

[2]We logged all points a registered user was served while position crafting.

system wanted the ranking to incorporate the ratio between views and votes [6].

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 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.[3]

[3]What 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

2.3 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 2.3). Many visitors opted to simply browse the guide, where they could see the pros and cons that others had already submitted.

Our results section draws on the following data sources, ordered by importance.

Activity traces. The timestamped activity traces captured in the database are our primary data source (Figure 2.3). 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

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.

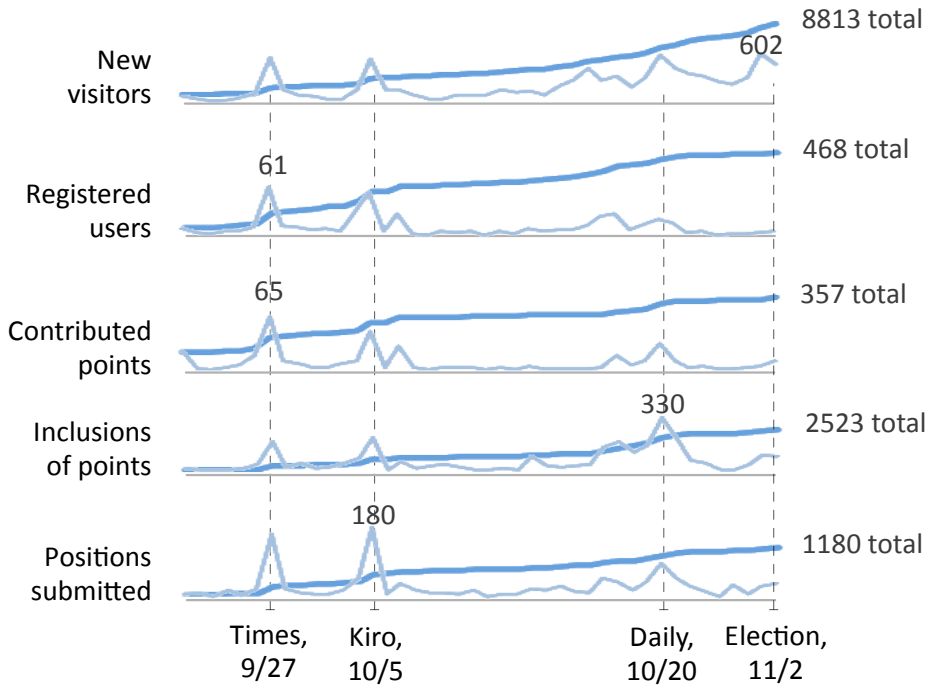


Figure 2.3: *High-level trends.* In every plot, each line operates at different scales: the dark line gives a cumulative count and the light line is the daily count. Highest daily peek and final counts are labeled.

(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.[4] 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.

2.4 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.

---

[4] A 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.

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**.[5] 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.

But inclusions and point authorship are lower bounds on the level of consideration of tradeoffs – users can still engage the arguments that others are making even if they do not result in pros and cons in their pro/con list. We can also examine the extent to which users explicitly requested access to see both pros on cons. We examine this from two data sources:

1. Point expansions. For points with a long description (45.4% of all points), users could click "read more" to expand the point and read the long description. This is evidence of wishing to access more information. For the 210 positions in which a registered user expanded at least one point on the respective measure, 39.5% did so for at least one pro and one con.[6]

2. Point pagination. When users were writing their pro and con list, the pros and cons that other people had submitted were displayed on the margins of the pro/con list, four points at a time. While the first four pro and con points were shown automatically on page load, all subsequent point listings were only served when the user explicitly clicked the "next" or "previous arrow. This activity thus

[5]We 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.

[6]We instrumented the site to capture all point expansions, though this data collection did not start until midway through the deployment (10/17). In this article, we use these expansions as indicators of whether users were attending primarily to one side of the issues or both. This data source is not unproblematic. On one hand, expansions do not necessarily indicate that the users actually read the points, and if they did, how attentively. On the other hand, the expansion data severely underestimates the amount of point reading that users may have done. First, as noted earlier, only 45.4% of all points even had a long description. Second, the 140 character summary was often quite well written, so users may often have not felt the need to read more, preferring to read additional points rather than more about a single point. We filtered the 3904 point expansions down to 714 by eliminating (1) expansions by anonymous users, (2) expansions by users who did not submit a position on the measure, and (3) duplicate expansions of the same point by the same user.

indicates users accessing pros or cons. For the 768 positions in which a registered user requested to view more points, 68.2% accessed both pros and cons.[7]

Overall, we believe this data indicates that LVG participants engaged in substantial consideration of tradeoffs. But 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.

Did the LVG facilitate engaging difference?

Balanced consideration is not just about considering tradeoffs, but also about whose points are being considered: did participants engage with users who had different views than themselves? In 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, or perhaps not even *read* their points. 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**

[7]We recorded every time that a point was served to a registered user in the position creation phase. Like point expansions, point views are not a precise measure of a user's reading activity. Point views were filtered by eliminating (1) the first four pro and four cons displayed unless the points were included in the user's position, (2) duplicate point views, and (3) point views by users who did not submit their position. This reduced the dataset from 29,508 to 10,609 views. On one hand, point views overestimate the amount of reading a user does: some users probably clicked through the points rapidly and simply skimmed the displayed points. On the other hand, point view data was only captured for points shown during the position creation phase, not on other pages where users could also scroll through points. Moreover, we did not capture point views by users before they logged in. Finally, we eliminated the first four pro and four con point views for each ballot measure a user viewed in order to account for points being automatically shown on page load. This is a conservative move because the first four pro and con points displayed are the points most likely to be read. Even if a user diligently read the full text of all of the first four pro and con points, we would not capture this in the point view dataset. Indeed, more than half of all point inclusions were of the points displayed automatically on page load. To mitigate this problem, we did not filter out point views of points that the user actually included into their position.

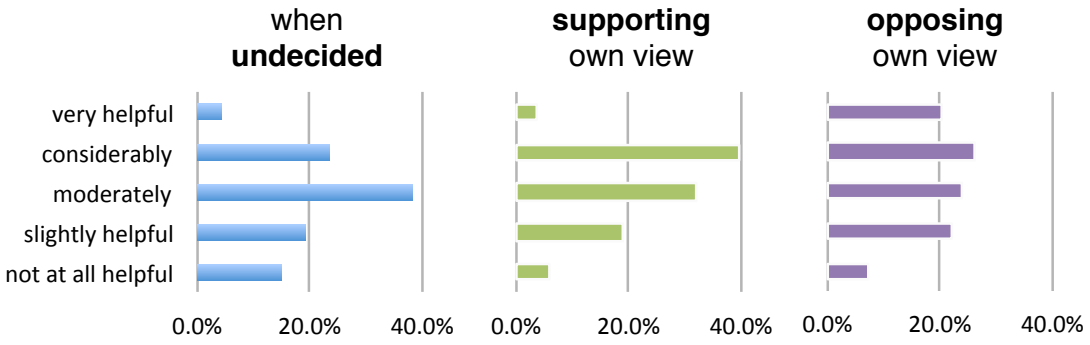


Figure 2.4: Helpfulness of reading points under different conditions, by fraction of survey respondents. Using pairwise 2-tailed Z-tests with $\alpha < .05$, we can reject the null hypothesis that the mean helpfulness of points is the same for those who are undecided vs. viewing opposing points, but 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.

point written by someone who took an opposing stance. Moreover, of the 165 positions where a non-neutral stance taking user expanded a point written by a non-neutral stance taking user, 56.4% clicked "read more" on at least one point written by someone who took the opposite stance.

We suspect that the prevalence of users including (or reading) 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 [82] 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, $r^2 = .249$, $p < 10^{-12}$). Similarily, 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 some insight. We asked three relevant questions. (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 2.4. 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 helpfulness 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 [76] term). 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 [76]).

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 assess, 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. A possible confound is that a user may have talked to someone outside of the LVG who influenced their opinion. Still, 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.

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 moderated 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 drilldown 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.

2.5 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 though 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 [2, 54]. 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 [104] by not making a value judgment about the validity of the source—anyone could decide for themselves whether a link to a pundit's blog is a reliable source or not. In our 2011 deployment, we enabled point authors to link to sources, but the functionality was used minimally. One future direction would be to enable other users to put in a request to the author to find a source, as this is a common request that commenters make on discussions around points.

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 [31]. 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 [31] and social identity theory [82] 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 [39]. 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 2.3).

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 [82].

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 others' thoughts 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.

Rephrasing is particularly interesting, as it implements the Reflect technique in a different context. For example, rephrased points could be available in the extended view of a point. Moreover, having multiple rephrasings of the same point can be useful for (1) summarizing and (2) clustering points.

2.6 Living Voters Guide, 2011

In 2011, we deployed a redesigned Living Voters Guide. The election was an off-off year (no presidential, no congressional), and the ballot measures were not as contentious as in 2010. Still, the 2011 Living Voters Guide saw increased visitation. Nearly twice as many unique visitors came to the LVG in 2011. The shift came from search engines. In particular, 50% of our traffic came from Google ads. Direct traffic and referral traffic (e.g. from a link from a news story or Facebook) was lower. Visitors were more widely distributed from across Washington than in 2010. Average time on site was down: over five minutes in 2010 to 3.5 in 2011. The increased traffic did not translate into more registrations and contributions than in 2010. There were 150 total pros and cons contributed, ignoring the 18 seed points, which is down from 360. 352 users created accounts, down from nearly 500. While the number of positions submitted were roughly equivalent (1043 vs. 1180), the number of point inclusions was down drastically from 2523 to 1321.

Aside from many cosmetic enhancements, our primary changes to the platform included:

1. Supported local ballot measures alongside state measures. Participants could enter their zip codes to unlock their local issues. 66.5% of all registered users elected to do so. 21.9% of all submitted positions were about local measures.

2. Pro and con points were more elaborate. First, users could add links to external resources when first writing their pro or con point. However, few people did (only 4.1% of points included a link). Second, users could expand a point to read the details as well as participate in a discussion rooted to the point. This was extensively used: there were 165 comments on the 169 total points.

3. Users could create their own pro/con lists and add their own pro/con points without creating an account. However, if they wished to save their positions or comment on a point, they had to create an account.

Overall, we saw an increase in traffic, less on-site contributions, particularly a reduction in the average number of points included per position, but saw a great increase in the amount of conversation around points. We have yet to analyze the 2011 data in terms of the metrics presented earlier.

2.7 Conclusion

Our primary contribution presented in this chapter is a novel extension of 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.

Chapter 3

PROMOTING LISTENING ON THE WEB WITH REFLECT

3.1 Introduction

As discussed in earlier chapters, to accomplish collective goals, people need to discuss what to do, how to do it, and who will work on what, whether it is an open source project, a faculty meeting, or a government planning process. At the same time that people increasingly expect deeper relationships with organizations [9], organizations are exploring new ways of including their constituencies into decision making processes. The input can be merely consultative, as with Facebook's initiative to gather feedback on terms of service changes and U.S. President Obama's participatory gathering of ideas for open governance. Or it can be at the heart of organizational culture, such as in Wikipedia where participants strive for wide community consensus in all decisions [10, 45]. We believe that these deliberative efforts to tap the creative energy, emotion, and intelligence of many are needed to improve our collective ability to confront public problems.

Unfortunately, public discourse on the web is often inflammatory and hyperbolic, or at least perceived so [27], which likely constrains the potential of web-based public discussion for informing decisions. We hypothesize that constructive web discussion is elusive and fragile partially as a result of the lack of attention paid to supporting listening. With attention to listening, interfaces might nudge people to establish more common understanding, even when the discussion is heated and listening is often forgotten. When participants establish some common understanding of a speaker's message, they are more likely to recognize the legitimacy of that person's perspective [81], and the quality of their opinions are often improved [16]. By returning to the basics of communication and developing an applied understanding of the role of listening in discourse we might be able to address a constellation of problems confronting participatory governance efforts. Our goal in this chapter is to draw attention to the generative potential of a deeper understanding of listening, and illustrate it through the design, deployment, and evaluation of a novel interface.

Our first contribution presented in this chapter is a design-oriented theoretical frame for supporting listening in web interfaces. Our second, and primary, contribution is a novel interface, Reflect, which illustrates just one point in the design space opened up by the listening frame. Reflect introduces a second dimension into online comment boards, adding a backchannel for people to demonstrate evidence of listening by restating the points that they hear the commenter making. Inspired by Wikipedia, Reflect nudges people to strive toward a neutral reflection of what others are saying.

Even if people cannot agree, perhaps some degree of consensus can be found about what is being expressed. Our third contribution is a presentation of Reflect's deployment in three diverse contexts: a high-volume technology news discussion site, summarization by facilitators during Wikimedia's Strategic Planning Initiative and during a local civic engagement effort, and finally an A/B tested discussion about Mechanical Turk on Mechanical Turk. Just as Reflect illustrates only one point in the design space, our analysis of these deployments does not test all of the theoretical propositions, nor all of the potential affordances of Reflect. However, we are able to demonstrate that restatement is an activity in which web users will voluntarily engage.

3.2 Interfaces for Grounding and Listening

Listening is integral to communication and it is not passive. Listeners must provide evidence to speakers to show that they are being heard and understood [19]. Listeners operate a backchannel where they nod, say "uh huh", tilt their heads, finish sentences, and preface their responses with a simple restatement of what the speaker said [30]. These explicit acts of listening helps speakers "debug" their messages, as well as provide evidence that they are being recognized and heard [101]. It also helps listeners demonstrate good faith as conversation partners; partners who provide more feedback through the backchannel are perceived as more patient, polite, and attentive [105]. The process of turn-taking, speaking and listening, exchanging evidence and repairing breakdowns, is called *grounding* [19].

In contrast, web interfaces have typically not explicitly supported backchannels for others to demonstrate evidence of understanding. Consider two ends of the spectrum of current (implicit) support for listening:

1. Threaded forums allow replies, which can be appropriated by users to listen, such as by rephrasing what is being said or asking for clarification. But mixing replies and acts of listening can quickly make a discussion difficult to follow if many people are involved. Moreover, there is an implicit expectation that comments add something new to the conversation. In some limited situations, the discussion format can clarify when listening is occurring, such as with the Q/A format.

2. Many discussion systems allow participants to vote on their favorite utterances, such as with

"like", "recommend", or "thumbs up" buttons. These buttons overload two functionalities: first, they perform an important listening functionality, in that they signal to the commenter that they have been heard; second, they serve to *judge* the statement. These mechanisms are not pure listening mechanisms, but also contain within themselves an implicit "+1, I agree" reply. This type of engagement thus does not serve those who may not in fact agree with what is being said but none-the-less recognize and appreciate the speaker.

Overall, web interfaces implicitly privilege speaking over listening, creating a feedback chasm that may hinder formation of common ground. Prioritizing speaking over listening may bias toward emotionally-charged but shallow interactions, which, in turn, likely affects who is willing to participate. We argue that interfaces can be designed that help nudge people toward more reflective interactions by emphasizing the common experience of listening. We hypothesize that:

1. **Listening interfaces enable people to see evidence that they are being heard, improving their communication satisfaction and willingness to participate further.** Consider reactions to the frustration of not being heard. One option is forcefully restating a point, further deteriorating the situation. A second reaction is to disengage. Being responded to is one determinant of whether people continue contributing in an online forum [5], and a similar dynamic may exist for perceiving that one is being heard.

2. **Listening interfaces empower participants to use and hone their active listening skills to guide the discussion and demonstrate their own worth.** Acts of listening impact the direction that a conversation takes [101, 19], and help to demonstrate to others that the listener is a valued participant. Consider the motivations for active listening: first, a listener might try to demonstrate understanding before responding (such as by restating), a commonly recommended technique for effectively interacting with someone about a controversial or complex issue (e.g. marriage counseling or dispute mediation in Wikipedia [11]). Second, listeners might provide feedback that helps teach speakers how to better frame their points. Third, listening can draw other people's attention to a statement's importance (e.g. if the point reflects their own personal beliefs [27]). Fourth, someone might articulate understanding in order to better interpret the intended meaning.

3. **By showing active listeners' demonstrations of understanding (or misunderstanding), listening interfaces help other discussants make sense of what is being said and why.** In an examination of how speakers and listeners work together to coordinate a conversation, Kraut found that while an active listener is the one whose understanding of what a speaker is saying is most positively affected when interacting with the speaker, eavesdroppers also benefit [59]. This effect is familiar: consider a twenty-person meeting where your colleague succinctly rephrases your overly long and convoluted expression of an important point. Your colleague has demonstrated that she has understood, and, in doing so, also helps everyone else understand better. These are acts of facilitation that lead others to say, "Ahhh, I see!" Listening interfaces may help elicit, capture, and expose these "ahh" moments for all future participants.

4. **Listening interfaces help establish an empathetic normative environment.** Even simple changes to a commenting interface can impact the thoughtfulness of comments, leading to the establishment of commenting norms as participants calibrate to the observed behaviors of prior participants [93]. Recent research in neuroscience indicates that the likelihood of engaging in empathetic activities such as perspective taking is sensitive to experimental manipulation [66], suggesting that we might successfully embed cues for empathetic listening in our interfaces. If the interface can encourage some users to listen, perhaps others may follow, helping to establish constructive communicative norms. Through their visible structuring, web interfaces can continually remind people to actively listen, even when the discussion is heated, and re-present these acts to other participants to bolster the listener's conversational status.

We have not tested all of these hypotheses. Rather, we contribute the speaking/listening theoretical frame as a useful metaphor for considering new design directions and guiding future empirical investigations, and demonstrate its productivity with the interface presented next.

But it is still just a metaphor. There are significant differences between the asynchronous character of the web discussions we target and the synchronous dyadic face-to-face communication upon which much of the research cited thus far is based. For example, we know that the introduction of

asynchrony, even short delays, can disrupt speaker/listener coordination [102]. Some may argue that reading/writing is a more accurate metaphor for asynchronous web discussions. While technically true, the reading/writing metaphor only indirectly captures the communicative functionality that listening highlights: timely, fine-grained feedback for communicators. Reading implies interaction between a reader and text, not interaction with the writer, whereas writing encompasses both raising new points and addressing points raised by the other. We find listening to simply be a fresh metaphor for imagining new design directions rather than the publishing frame implied by reading/writing.

There are a range of listening mechanisms that can be designed to engage the themes described in this section. For example, a minimal listening mechanism could be explored, a simple "I hear you" button. Wikimedia in their Wikilove campaign has recently started to discuss including a "thank you" button in place of a "like" button, which would operate similarly. At the next level, a non-textual listening tool for supporting grounding could be designed, where users can select text and annotate it with an emoticon that conveys their sentiment when reading that piece of text (e.g. smiling, grimace, quizzical). These emoticons from multiple people could be aggregated and visualized on mousehover, helping a commenter debug their message. Taking the listening metaphor to its furtherest extent would be to facilitate users in restating the points that they hear the commenter saying. This nudge to restate is motivated by a useful practice when in an emotional argument: restating the other's points before responding.

In this chapter, we experiment with this latter strategy of restatement. By carving out a space for restatements in a comment board, we gain a great deal of leverage for inquiring about what would happen if we make active listening a first-class activity in online comment boards, helping commenters know how their statements are being understood, focusing others on slowing down to understand what is being said, and providing fodder for later summarization activities.[1]

---

[1]I am often asked "why restatement, as opposed to a more minimal listening interface?". In other words, why Reflect? The answer is that there was no actual choice made. Every aspect of the framing presented here is reflective, in that it happened after I had a eureka moment that rephrasing incivil comments was a potentially high-value line of attack on counteracting some of the problems with comment boards. After months of designing and iterating on this intuition, I started to understand that the fundamental concept that I was working with was listening, and still more months later did I come upon the literature on grounding. Therefore, by the time I started to frame this work under the rubric of listening and grounding, the choice of how to facilitate it was already fixed on restatements. This is a bottom-up strategy for design research: focusing on solving real problems and following one's intuitions, trusting that there are intellectual concepts worthy to be articulated that will eventually be uncovered, and which will in turn then suggest additional design opportunities. I strongly believe this bottom-up approach is just as legitimate as a top down approach of abstractly articulating a high-level design space first and then enumerating different dimensions, and deliberating

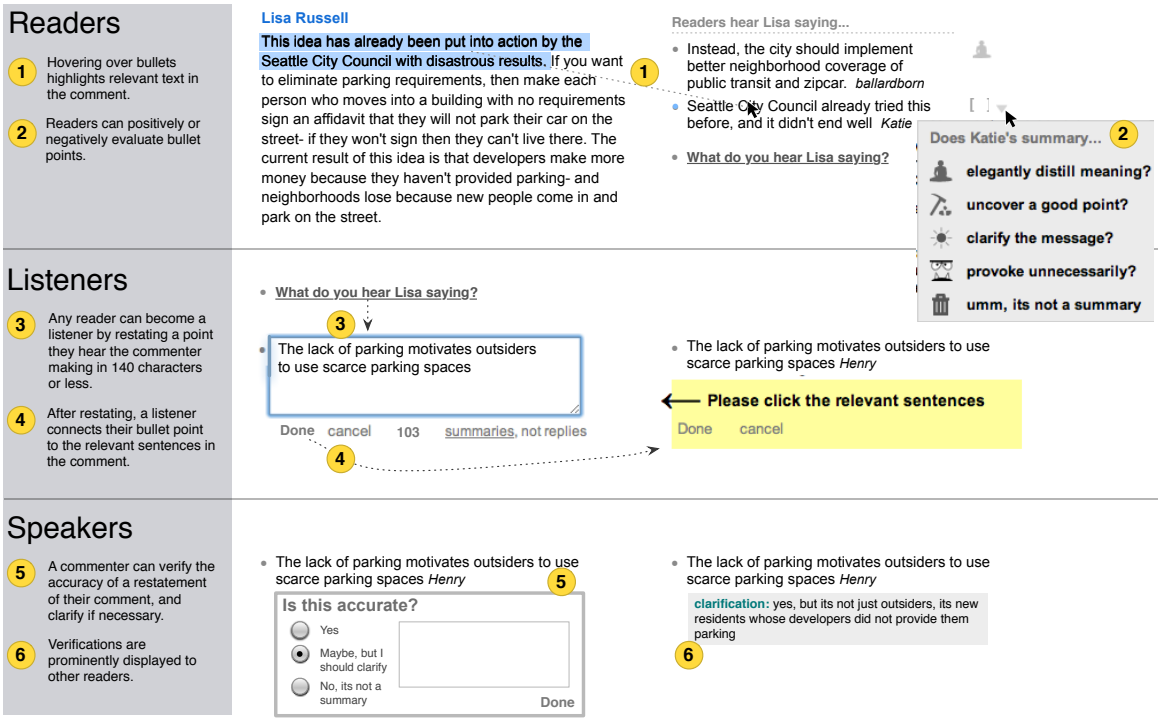


Figure 3.1: Mechanics of the Reflect interface.

3.3 Reflect: Listening Through Restatement

In this section, we present a novel interface that explores a web commenting experience strongly situated in the listening metaphor. Reflect makes a lightweight change to online comment sections: to the right of every comment, visitors are invited to restate points they hear the commenters making (Figure 3.1). These restatements, each an act of listening, are publicly displayed in a bulleted list. Reflect provides an opportunity for the original commenter to respond to each bullet point to verify its accuracy and clarify if necessary.

Reflect thus establishes a listener's backchannel that engages the theoretical propositions laid out earlier. First, commenters can learn how others are interpreting their statements and clarify misunderstandings, supporting one iteration of grounding. Second, listeners can gain status as pro-

about which option along some spectrum would provide the most intellectual leverage. If I had not followed my intuition, I would not have discovered this conceptual framing of the design space around listening in the first place.

ductive listeners. Third, other participants can scan the bullets to better understand what is being said and identify important points, assuming that listeners are more likely to restate points that they find important (a form of "read wear" [46]). And finally, in contrast to most comment boards, a Reflect-enabled site strongly suggests to a visitor that listening is also a primary mode of participation, perhaps helping to establish a normative environment through which common ground is more likely to be found.

3.3.1 Scenario of use

Imagine that you have been participating on the Ideas for Seattle website, where residents suggest ideas for improving local life, and vote and comment on them. You submitted an idea recommending the elimination of an ordinance requiring real estate developers to provide for future tenant parking on the grounds that the requirement is too rooted in the assumption that city dwellers need a car. A few hours later, you check out the comments your idea has provoked.

You start browsing the comments, glancing over the first one and then read the summary points others have written. When you hover over a summary bullet, the relevant text in the comment that the bullet refers to is highlighted, allowing you to jump between the commenter's own words and the listener's interpretation (Fig. 3.1.1). This particular summary is interesting because it helps you better understand the point the commenter was making about the free-market implications of your idea. You mark the restatement as "Helps shed light on what the commenter was trying to say" (Fig. 3.1.2).

You recognize the next commenter, Lisa. She was the one who added the helpful restatement. You note that no one has restated any of her points, so you read her comment carefully. You decide to add a bullet point so that she knows someone is listening and so that her comment is not lost in the mix. After clicking "Add a point that Lisa made", you type in your interpretation of her main argument in the 140-character limited text field (Fig. 3.1.3). It causes you to stop and think for a while, which is often not how you usually skim comments. After hitting submit, you are asked to connect the point that you summarized to the part of Lisa's comment to which it refers (Fig. 3.1.4). After clicking two sentences, which are subsequently highlighted, you click done. The bullet shows up in the list.

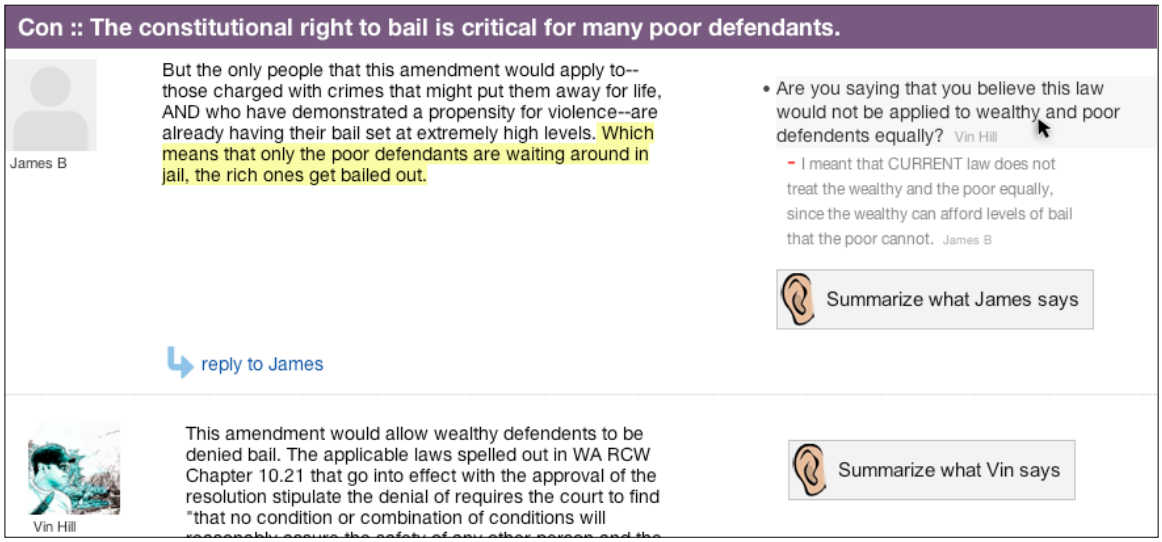


Figure 3.2: The first use of Reflect in the wild, on the Living Voters Guide. In this exchange, Vin first restates a point he thought James was making, and then responds to James' comment. James received an system-generated email indicating that someone restated a point he made. He followed the e-mail's link and clarified his point to correct Vin's misinterpretation. The exchange illustrates a major issue with asynchronous listening: Vin replied to James' comment before James came back to verify and correct Vin's restatement of his point. On one hand, anyone browsing the forum after James responded to Vin's summary would benefit by avoiding Vin's misinterpretation. But Vin's own reply now makes less sense.

Scrolling down, you see a comment you had left earlier. Several people have added bullet points. You recall having received a notification email earlier, but had not had time to check it out then. The interface presents you with the option of verifying the accuracy of each bullet point, or if it is even a restatement at all. You can also add a short clarification if needed. The first bullet is accurate. You click yes and a checkmark appears next to the summary. The second bullet is inaccurate. You click "no" and clarify what you were trying to say (Fig. 3.1.5). The clarification shows up underneath the bullet point (Fig. 3.1.6).

3.3.2 Design rationale

Here we describe five important Reflect design choices.

1. **Articulating the work of comment summarization.** There are other ways in which we could break up the task of restating or summarizing the comment besides short bullet points. For example, our first design used a wiki next to the comment. However, in early testing, this proved problematic. People felt that with a wiki, they must commit to summarizing the whole comment (many of which can be quite long). We moved to bullet lists and discovered some additional affordances: (1) we could more naturally include more people in the listening and grounding activity, (2) enhanced support for clarifying singular points, (3) facilitating multiple interpretations of the same text, even at different levels of meaning, (4) easier for commenters to respond to the summary and clarify their points, (5) easier to skim the summaries and connect them to the comments, (6) the ability to identify similar points being made in other comments, and (7) the ability for listeners to highlight the points they found most salient in what the commenter said.

2. **Connecting summary text to comment text.** After someone summarizes a point that a commenter makes, we ask him or her to click on the relevant sentences in the original comment where the point was being made. There were two reasons for doing this. First, accountability: a listener explicitly commits to exactly where a commenter made the point they believe them to be making. For those who are maliciously misinterpreting the message or deliberately antagonizing the commenter, it adds a step whereby they are forced to essentially acknowledge their maliciousness. Second, readability: this data is used later to aide readers in figuring out the relevant parts of the comment to which the summary refers. If there is a breakdown in grounding between speaker and listener, the speaker may be able to more easily see where the misinterpretation occurred. Others can judge whether the listener accurately represented the point. If someone maliciously writes a bullet, it's relatively easy to verify what that the sentences actually said.

3. **Speakers are privileged to verify summaries.** Because Reflect places the bullet points next to a comment, an inaccurate bullet might overly influence what others believe the commenter

was saying. Therefore, Reflect enables speakers to verify a summary bullet point and clarify when needed, which is prominently displayed to other readers underneath the bullet point. However, Reflect does not entrust speakers with the ability to directly edit or delete the bullet points. Otherwise they may overly censor other people's valid, read-between-the-line interpretations. Privileged responses address the tension between allowing the speaker to repair breakdowns with giving listeners the power to express their interpretations. These verifications thus facilitate public grounding of expressed meaning.

4. **Community evaluation of restatements.** Not every restatement is created equal. Some restatements might highlight a point deeply buried in a long comment, or distill the essence of a long comment. Or bullets might be used to create snarky replies or offensive, off-topic content. For these reasons, we wanted to enable other participants to recognize good restatements and to flag poor ones. Aside from potentially removing malicious restatements, this can help to articulate for participants the purpose that good restatements might play. Note that this functionality was added during later design iterations and was not present in the Wikimedia and CCF deployments discussed later.

3.3.3 Implementation

Reflect is implemented as a jQuery-based Javascript library, available under the MIT open source license. This library requires a configuration object specifying CSS selectors describing the DOM for the comment section. On page load, the core library can then wrap the Reflect markup around the relevant elements. In order to embed Reflect, a server-side web service for fetching and storing Reflect data must be implemented. The service can also carry out other tasks like email notifications. Full Reflect implementations are available (1) for the two most common content management systems, Wordpress and Drupal; (2) as plugins for the Mediawiki and Slash platforms; and (3) for the Rails and Django web frameworks.

3.3.4 Related design work

By encouraging participants to seek mutual understanding of each others' comments, Reflect supports the process of grounding described by Clark's contribution theory of common ground [19].

Reflect, to our knowledge, is the first interface that makes the process of content grounding a first-class activity in web-based discussions. While Clark's theory has been used widely in HCI research, it has typically been used an analytical lens to understand the affordances of the communication medium (e.g. [38]) or how communication strategies change under pressures of scale [103]. Systems that are designed specifically to support grounding tend to be for distributed teams coordinating their activities [24].

Other systems afford restatements, such as annotation and anchoring tools [13, 18], but do not call out listening functionality explicitly. The only work we have found that explicitly supports restatement is ThoughtSwap, a domain-specific classroom tool for supporting co-located anonymous discussion [28]. Students submit ideas in response to a prompt, and then other kids can pull ideas out of the "hat" and re-present those ideas.

Other interfaces have been designed to help users navigate threaded discussions, such as interfaces for visualizing the social behaviors and semantic associations [84, 91], or focus+context interfaces for reading threads [98]. OpinionSpace [32] 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. These interfaces present novel methods for aiding users in navigating discussion spaces. They can help set the context for listening, but they do not go beyond the typical rating of comments when it comes to engaging with what is being said.

Ultimately, Reflect is meant to incite deeper reflection about what others are saying. In this sense, one of the most related systems is MetaViz [7]. MetaViz attempts to trigger critical thinking by using computational metaphor identification to expose the political metaphors that participants are drawing upon in a blog. The evaluation of MetaViz demonstrated promising patterns of reflection and creative thinking, even when the algorithm contributed an erroneous metaphor.

3.4 Field Deployments

To provide insight into how Reflect might be appropriated in real discussions, we have deployed Reflect in four distinct situations: (1) during four stories posted to the popular technology news

discussion site Slashdot; (2) a discussion between Mechanical Turk users about how Amazon could improve labor conditions on Mechanical Turk; (3) Reflect's use by the Wikimedia Foundation to summarize input they received during a worldwide Strategic Planning Initiative; and (4) as a means by which the leader of a local civic engagement initiative (CCF) could show participants on their blog that he was listening to them.

We divide our deployments into two contexts of use: *community news discussion* (Slashdot) and *community deliberation* (Mechanical Turk, Wikimedia, CCF). Because of the size of Slashdot's readership, the Slashdot deployment helps us most in establishing whether Reflect will be likely to be used by a wide public, and the extent to which Reflect would be used to listen. The three deployments in the participatory governance context are more limited because details of each deployment limited the size of the user base. However, the value of these other deployments is to illustrate different possible use cases of Reflect, and listening interfaces more broadly, for supporting aspects of participatory governance. All of our deployment contacts and interview participants gave us explicit permission to use their real names.

3.4.1 Community News Discussion

Two fundamental questions about Reflect is whether online discussants will (1) actually use the Reflect functionality and (2) whether they will use it to listen to each other, rather than troll or reply to commenters. These questions are essential in determining if Reflect, and listening interfaces more broadly, might have traction for broad public use. If it does, then pursuing this line of inquiry further may be fruitful, both in terms of creating new designs for listening as well as rigorously examining if the hypothesized outcomes laid out earlier about changes in interpersonal behaviors are realized.

To answer these questions, we deployed Reflect on four articles posted to Slashdot, one of the most popular news discussion sites since its founding in 1997. Slashdot's audience is primarily male professional engineers, an audience one may not predict to embrace Reflect and its emphasis on empathetic listening. We argue that if the Slashdot community uses Reflect as intended, it indicates that Reflect may be widely applicable. However, there is a limitation to this argument: the skilled user base may be more capable of understanding Reflect than the general public. Moreover, the Slashdot platform itself, with its meta-moderation and reputation system [67], has been one of

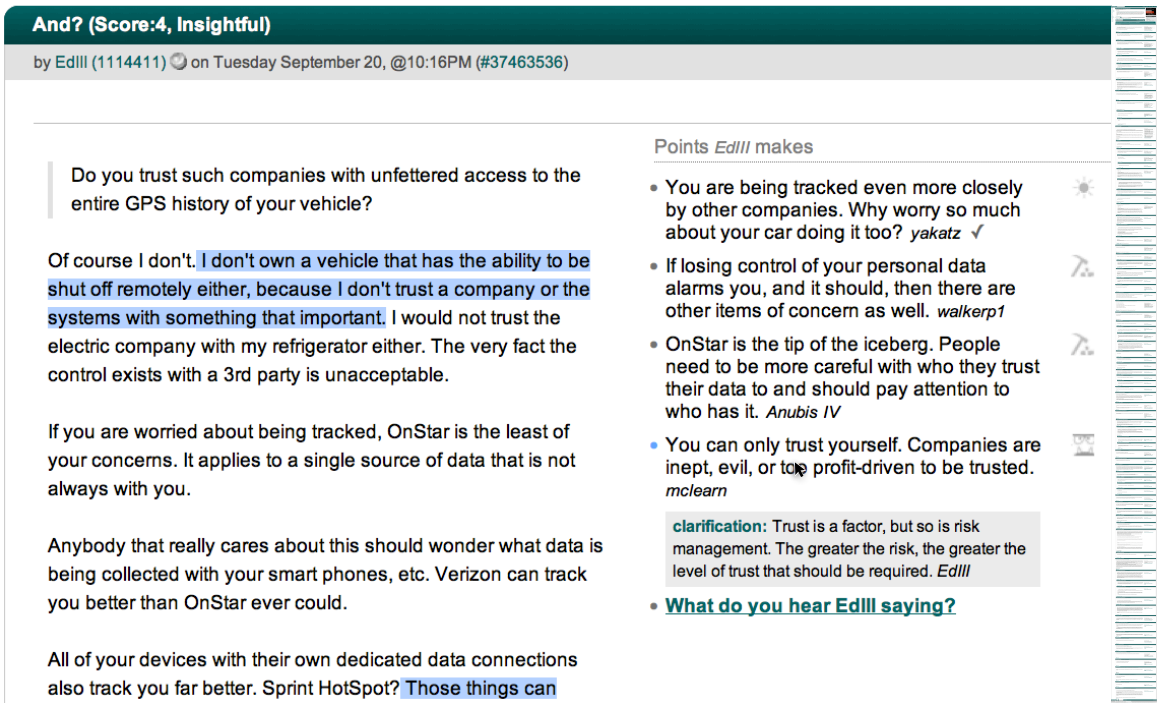


Figure 3.3: An excerpt from a Reflect-enabled Slashdot story. The entire comment section is embedded to the right in the margin. Readers of the digital version can zoom in to peruse the whole discussion. Those who printed the pdf will need to fetch their magnifying glasses.

the pioneers in web discussion technology and so the Slashdot user base is already familiar with complex commenting systems.

The opportunity to deploy Reflect on Slashdot was made possible by an introduction by fellow researcher Cliff Lampe, who introduced me to Slashdot founder Rob Malda. From Spring to Fall 2011, we developed the Slashcode plugin and coordinated with Malda and engineer Tim Vroom. Through this process, Reflect's design evolved, including the addition of the community moderation functionality. This emerged as an important design addition because Malda, with his decade-and-a-half experience moderating online discussions, was concerned that Reflect would be used to troll the commenters or to reply to commenters rather than summarize. We designed the bullet evaluation functionality, and then used the evaluations to deactivate bullet points that were rated negatively. Specifically, in our scheme, if more than half of the evaluations of a bullet point were negative after

Table 3.1: Absolute counts of four contribution types in the Slashdot data (and number of unique users contributing each).

	Comments	Bullets	Bullet ratings	Bullet replies
Story 1	389 (145)	353 (113)	719 (285)	20 (13)
Story 2	112 (67)	152 (47)	335 (203)	17 (7)
Story 3	57 (37)	84 (36)	181 (129)	14 (6)
Story 4	174 (93)	169 (65)	264 (144)	12 (8)

the first three evaluations, the bullet was deactivated and hidden. The bullet was also deactivated if the commenter labeled the bullet as "not a summary" during verification and at least one reader gave a negative rating.

In mid-September 2011, Slashdot enabled Reflect on four stories. During the trial, **773 bullets were written by 247 discussants, an average of 1.1 bullets per comment** (Table 3.1).[2] Figure 3.3 shows an excerpt of Story 4. The full discussion is embedded in the margin.

Who is listening?

15.3% of commenters in the stories also created a bullet point, overall contributing 41.2% of the bullets. The remaining 58.8% of the bullets were contributed by who we may previously have labeled as "lurkers". Reflect may be encouraging participation from those who do not usually comment.

Are commenters verifying the accuracy of bullet points?

Slashdot users verified 12.8% of bullet points. This may have been higher if commenters would have been notified by email about restatements, but would have created a deluge of email. Of the verifications, 60.6% characterized the bullet as not a restatement, 8.1% were clarifications, and the remainder confirmed the bullet's accuracy.

[2]This excludes 375 postings of the same bullet by a single user operating four accounts. Despite participation by 230 users in submitting 977 bullet ratings moderating this user's bullets, the Slashdot engineer still needed to run a query to deactivate this user's bullets. Clearly, more work needs to be done to better handle this scenario. In our analysis, we have excluded from our dataset all data related to this user's activities.

Did the written bullet points actually reflect the meaning conveyed in the comment?

Our most important empirical results, after establishing the baseline that Reflect would in fact be used, regard characterizing the extent to which bullet points actually reflected back the commenter's points, as opposed to simply being replies, personal attacks, or off-topic posts. And if Reflect was in fact used for listening, were there different ways in which bullets reflected meaning? To answer these questions, we performed content analysis on the bullet points. This was a challenging task because the bullet points turned out to be deeply contextual, rich social expressions drawing on multiple levels of shared meaning: the Slashdot culture and past interactions; implicit references to the story being discussed and the comments and commenters that preceded the comment under question; and reflection upon fairly detailed, nuanced comments.

Coding methodology We coded all 353 bullets created for comments about one of the stories using a standard grounded analysis through which we iteratively developed and refined a codebook and establishing inter-rater reliability. Ultimately, we decided to have each coder independently code every bullet and resolve discrepancies in order to yield results in which we are confident, even on the low-level codes that we did not achieve satisfactory reliability for independent coding. See the following footnote for details.[3]

The codebook has two levels of codes, shown in Table 3.2. The highest level is a yes/no code: Does the bullet primarily reflect back material from the comment? We achieved $\alpha = .89$ reliability for answering this question. The low level codes regarded classifying a bullet point after it was determined whether it reflected meaning or not.

The codebook has five primary subcodes for bullet points that reflected meaning. First, *pure reflections* were neutral restatements that rephrase, distill, or elaborate upon (e.g. through exam-

[3]The codebook was developed by categorizing a random sample of 50 bullets, then iterated on with Jonathan Morgan and Alan Borning as we independently coded and resolved discrepancies on a few sets of additional randomly sampled sets of 25 bullets. We each independently coded 100 additional, distinct points each. After discussing the codebook's shortfalls in capturing much of the nuance, Jonathan Morgan and I developed a new two-level codebook (shown in Table 3.2), scrapping our previous codes and starting over. We managed to achieve high reliability on the high-level code but reliability on a couple of the subcodes was far less reliable. Our worst code, the tone code for "negative/cheeky" bullets, was $\alpha = 0.45$. Low reliability was due to difficulty identifying all the necessary cues on each code, not because we disagreed on phenomena; resolving discrepancies was in almost all cases fairly simple, as one of the coders generally saw something that had escaped the other's attention. We decided that we would each independently code every bullet point, and resolve discrepancies so that we would be confident in our codes, even for the subcodes.

Table 3.2: Codebook for classifying Reflect bullet points, as well as our reliability scores for suitability for independent coding (Krippendorff's Alpha) based on a set of 100 independently coded bullet points, and the overall prevalence of the code in our 300 coded bullet points.

Code	Reliability	Prevalence
Reflects back meaning of the comment	0.89	62.2%
Pure reflection that neutrally restates without elaboration	0.73	20.6%
Expands meaning by reading between the lines	0.66	34.1%
Meta observation that contextualizes the content	0.84	15.5%
Antagonistic or cheeky reflection of meaning	0.45	12.8%
Contains a reply in addition to reflecting back meaning	0.58	2.7%
Doesn't reflect meaning contained in the comment	0.89	31.4%
Flames the commenter with an ad hominem	0.84	15.9%
Off-topic and doesn't relate to the discussion	1.00	5.4%
Replies to the commenter without reflecting meaning	0.68	9.8%

ples) the respective comments' meaning. They do not go beyond the comment's meaning in any significant way. If *pure reflection* was selected, no other code could also be selected because the purpose of this code was to capture the extent to which Reflect was used for *strictly* its intended use. Second, the *expands meaning* code was used for bullet points that read between the lines, such as by generalizing, inferring antecedents or consequences, or putting a high-level label on a phenomena the commenter described (e.g. "Social justice" when the commenter discusses income tax equalities). Third, the *meta observation* code was applied to bullet points that contextualized the meaning of the comment, such as by focusing a critical eye on the rhetorical devices employed by the comment or positions the meaning with respect to the norms of the community or other external events. Fourth, the *negative/cheeky* code was applied to bullet points that are non-neutral, usually tongue-in-cheek sarcastic summaries that often mocked the ideas at play, or were overtly negative toward the comment's content. This code helps us distinguish bullets that, while reflecting meaning, do so in a potentially antagonizing manner. Fifth, the *contains replies* code was applied when

a bullet, after reflecting the meaning of the comment, also inserted a response to that point. This response was often contained within a second clause, with the first clause reflecting and the second replying. Finally, we also had a rarely used *other* category for bullets that did not fall into any of these subcodes, such as bullets that were written in response to other bullet points that served to correct the other bullet author's interpretation of what the commenter was saying.

There are only three subcodes for bullet points that do not reflect meaning. First, the *flames* code captures bullet points that simply served to personally attack the commenter via ad hominems. It does not engage the substance of the comment, but rather the commenter. Second, the *replies* code is for bullet points that reply to the comment without reflecting back any of its meaning. Third, the *offtopic* code is for bullets that do not appear to be related to the discussion in any way. Again, there is also a rarely used *other* category that was a catchall for non-reflects that did not fall into these subcodes, such as verbatim copying of a sentence in the comment.

Results To illustrate the rich range of bullet points users wrote and better demonstrate these codes, we now walk through a concrete example. One of the Slashdot stories was about the U.S. federal government closing down an online poker site because it was operating as a ponzi scheme. And one of the comments stated: *When Mitt Romney asks, "Why punish success?", I suggest people think about [poker sites like this]. Not all businesses are scams, but the people raking in millions of dollars a year aren't earning it. Their [sic] inheriting it, winning it or stealing it, and they deserve to be taxed at a higher rate."* This comment had more than 20 bullet points created for it, illustrating a range of these codes. Let's look at several of them:

1. *Just because someone obtained their wealth legally, does not mean it is wealth they have earned or generated.* Coded as **pure reflection**.

2. *Not all business is criminal, only successful business.* Coded as **cheeky, expands** because the bullet writer expands the meaning of the comment by equating "people raking in millions of dollars" with "successful business", in order to make it possible for the cheeky reductio ad absurdum.

3. *[the commenter] is emoting as per his training and will be rewarded by his fellow jealous*

and equally well trained primates. Coded as **flames** because this is an ad hominem attack that does not engage the comment. If, however, the bullet writer had said *[the commenter] is emoting as per his training about how the rich have cheated their way to wealth and should be taxed more, and will be rewarded by his fellow jealous and equally well trained primates.*, it would have coded as **cheeky, meta** because it reflects back meaning in a negative way, at the same as it makes a meta comment about how some other Slashdotters will respond to the comment.

4. *reflecting the current US topic of 'class warfare' related to taxing higher incomes at the same rate as lower incomes.* Coded as **meta** because it contextualizes the comment's point with respect to the larger social issue at play.

5. *Implication that Romney isn't so different from the people involved in the criminal scheme.* Coded as **expands** because it reads between the lines. As this example shows, bullets that expand meaning may be incorrect, as this inference does not definitively follow from the comment.

6. *eg: inheriting it, winning (don't know what that means), stealing: rent-seeking, legislating away competition, etc...* is a **pure reflection** because the bullet author is struggling to restate what they are hearing, staying within the bounds of the comment even as they elaborate it without expanding its meaning (*"rent-seeking, legislating away competition..."*). We did not code a bullet as "contains reply" for clauses such as "(don't know what that means)" because it is done in a manner that could not stand on its own as a reply.

7. *He thinks businesses are scams, and he thinks business people are stealing from him, while at least he gets entertainment in a casino.* This is a difficult one, but is coded **expands, cheeky, contains reply** because the comment is twisted such that the bullet writer can impute that businesses are personally stealing from the commenter, while the last clause about "getting entertainment" appears to be a humorous reply to the commenter about how the income inequality issue is worse than the casino operation.

8. *POLITICS LOL.* Coded **replies** because it is essentially calling out the commenter for turning the discussion toward politics, which goes against the Slashdot community's often reinforced norm regarding focusing on technical, not political, content.

Overall, **62.2% of the coded bullets Reflected back meaning in one form or another, indicating Reflect's use as a listening interface. Discounting negative/cheeky bullets, 49.3% of all bullets listened in an apparently neutral fashion.** Of the subcodes, "expands meaning" was the most prevalent, applied to 34.1% of all bullets. As coders who needed to also understand the comment, we found many of the bullets falling into this category to be the most useful because they often brought to light the implied meaning of a comment which was often unclear, subtly implied, or obfuscated because the meaning of the comment could only be understood with respect to a number of prior comments. While flaming and pure replies were fairly common, given concerns aired by Slashdot co-founder Rob Malda, we were pleasantly surprised that they were not more frequent.

The distribution of frequencies in Table 3.2 as well as the set of examples given above conveys a variety of rich ways in which Reflect was used to listen. In many cases, it appeared that Reflect users were in fact using it to help coordinate conversation, sometimes even being used as a normative device for cheekily deconstructing shared inside jokes, or punishing deviance (as with the "POLITICS LOL" bullet above).

Did the community's moderation of bullets align with the designed intent of Reflect by filtering out bullets that did not restate points?

During the Slashdot deployment, community moderation was enabled, as described earlier. When enough users rated a bullet as not a restatement of one form or another, the bullet was deactivated and hidden from view for subsequent users. Of course, the Slashdot community did not necessarily have to appropriate this functionality to negatively rate non-restatements; instead, they might rate down restatements, or there may be no difference in ratings for bullets that reflect and those that do not. An important question therefore is whether the community used the moderation functionality to enforce the Reflect functionality as a space for listening. To answer this question, we examine our coded bullets to determine whether the bullet points that were deactivated tended to reflect meaning or not.

We found that the community used the bullet evaluations to sanction and hide bullets that did not reflect meaning, increasing the visual salience of listening. 25.2% of all bullets were moderated by the community and hidden from future view. Of these moderated bullets, 71.6% were not summaries. This means that **after community moderation is factored in, 75.0% of the final set of publicly visible bullet points were legitimate summaries**. Moreover, 44% of the moderated bullets that reflected meaning were negative/cheeky. The 62.2% figure provided earlier thus conservatively states how prominent the listening functionality of Reflect was to a visitor to Slashdot. The community sanctioned non-restatements, carving out a backchannel for listening.

These results illustrate what discretionary use of Reflect might look like. We found an almost one to one comment to summary ratio. Moreover, many lurkers participated, potentially engaging a different motivational frame. Aside from raw use, we found that participants generally followed our design intentions in writing bullets that restated points, and that the community moderation further reinforced this intention. These results are particularly promising given the reputation of Slashdot as an often vitriolic community.

3.4.2 Community Deliberation

There are many scenarios in which a community needs to deliberate and sometimes make decisions about what to do about challenges the community faces. Sometimes this takes the form of a government, non-profit, or movement organization launching a project to engage its constituents around the topic. Or it might be a more bottom-up discussion started by a community member that influences decision-makers to take action (or foments opposition).

We see these efforts at community deliberation as essential components of *participatory governance*, and it is a type of discussion that we are most interested improving through listening interfaces. Here we describe three deployments of Reflect that each highlight distinct aspects of community deliberation that Reflect might be used to support. After we describe these deployments, we present interview data from the leaders of the latter two deployments. The empirical work presented in this section is preliminary and sometimes anecdotal. It is not conveying objective, generalizable knowledge. It is valuable, however, because it helps designers understand more deeply, if imperfectly, some of the different contexts of use to which Reflect might be put, and

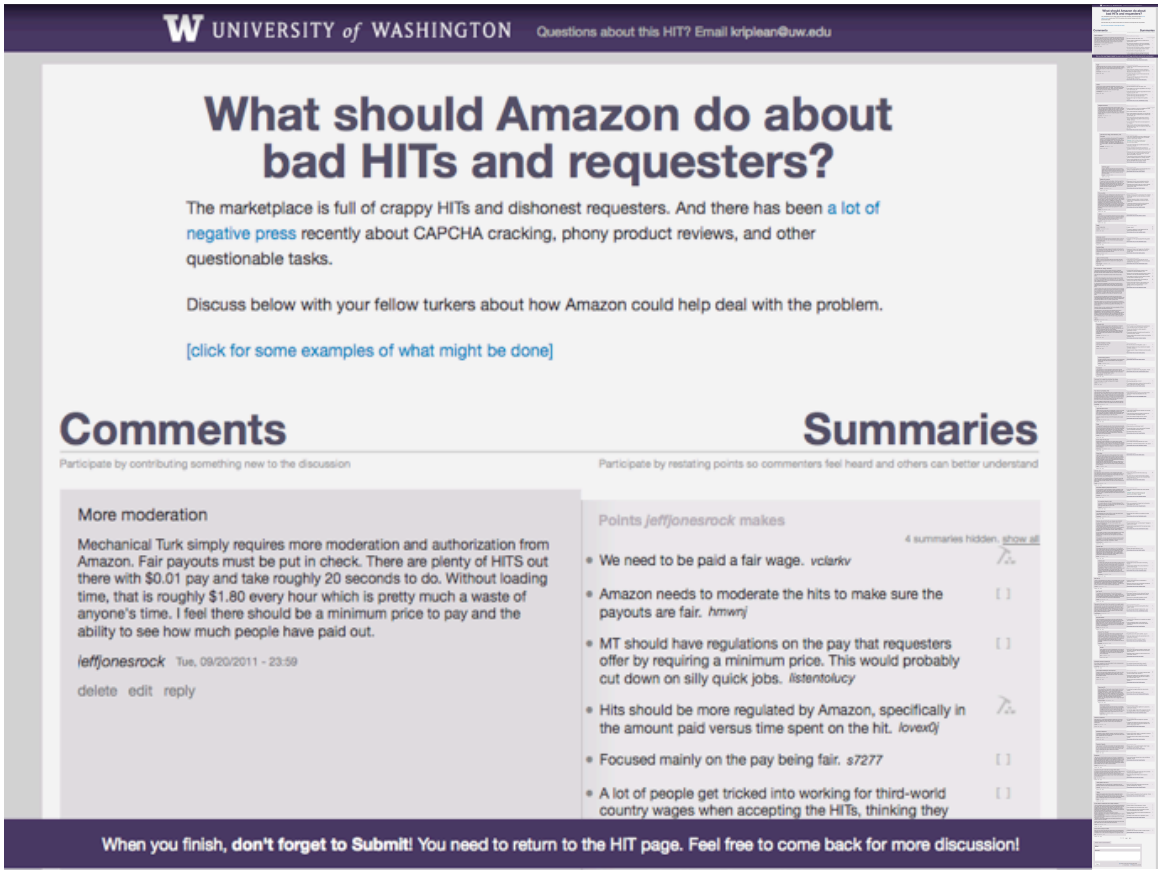


Figure 3.4: The discussion prompt for the Amazon Mechanical Turk discussion. LIke the Slashdot figure, a major portion of the Reflect-enabled discussion is embedded on the right-hand side of the screenshot. Zoom in on the digital version to read it.

highlights opportunities for future empirical work.

Case 1: Ideation in Mechanical Turk

Mechanical Turk is an Amazon web service where anyone can post small tasks to be completed for small amounts of money. People who post tasks are called "requesters", those who complete tasks are called "workers" or "turkers", and the tasks are called Human Intelligence Tasks, or "HITs". A community of workers has emerged, bonded by shared experiences doing these HITs, sometimes communicating on a website called "TurkerNation". One particularly high tension shared experi-

ence is being cheated by requesters who refuse to pay for completed work, as Amazon gives power to requesters to reject poor quality work, with few mechanisms for workers to appeal decisions. Another common issue are the many HITs posted to the Mechanical Turk marketplace that are part of efforts of dubious moral nature (e.g. cracking CAPTCAs or posting fake product reviews).

We decided to host a community deliberation about *"What should Amazon do about bad HITs [aka microtasks] and requesters?"*, using a standard Drupal-based commenting board. We invited Mechanical Turkers to participate in the discussion by posting a HIT with a discussion prompt, invitation to join the discussion, and statement of our intent to report the outcomes of the discussion to Amazon to the best of our ability. Participants were paid for completing an intro survey that they filled out before being redirected to the discussion, and not for their participation in the discussion, though afterward, we informally rewarded bonuses for good participation. The discussion was A/B tested: there were two simultaneous, independent discussions we hosted. After participants "accepted" the HIT, they were randomly assigned and redirected to one or the other (based on a hash of their worker ID). To control for the heavy influence of initial comments on the tenor of a discussion [93], we seeded both discussions with the same three initial comments, drawn from a pilot run. Participants were also asked to fill out an exit survey. The survey was an attempt to evaluate whether there might be major changes between the two conditions in how people perceived other participants and the discussion more broadly. See Table 3.4 for the questions. By conducting a deployment on Mechanical Turk, we gain an ecologically valid naturalistic deployment (participants discussed an issue they cared about, were knowledgeable about, and were not directly compensated for their participation), while still garnering some level of experimental control through the independent discussions. Figure 3.4 shows the discussion prompt and an except from the Reflect-enabled discussion condition.

Behavioral results We recruited over a hundred participants. Many workers decided to participate in the optional discussion in both of the conditions: 95 comments (45 replies) were added in the Reflect condition and 77 comments (28 replies) in the non-Reflect condition. Based on our informal evaluation, there were few offtopic comments, with most appearing to sincerely engage the topic.

In the Reflect condition, 142 summary bullet points were created (1.49 bullets per comment). Using the same codebook and methodology developed for coding the bullet points from the Slash-

Table 3.3: Bullet point classification for points created in the Reflect condition of the Mechanical Turk deployment. Adopted the same codebook and used the same methodology as with Slashdot.

Code	Prevalence
Reflects back meaning of the comment	92.0%
Pure reflection that neutrally restates without elaboration	65.0%
Expands meaning by reading between the lines	26.0%
Meta observation that contextualizes the content	1.0%
Antagonistic or cheeky reflection of meaning	1.0%
Contains a reply in addition to reflecting back meaning	0.0%
Doesn't reflect meaning contained in the comment	8.0%
Flames the commenter with an ad hominem	0.0%
Off-topic and doesn't relate to the discussion	0.0%
Replies to the commenter without reflecting meaning	8.0%

dot deployment, we coded 100 of these bullet points. The results are shown in Table 3.3. The bullet points were almost exclusively pure reflections or expands meaning, with a few bullet points that replied to the comment rather than reflecting meaning. There were no offtopic bullets, nor any flames. Overall, the social dynamics and listening practices revealed in the Slashdot deployment were far more complex and nuanced; the reflections in the Mechanical Turk deployment were straightforward.

Survey results Our preliminary results from the exit survey in this first experiment do not show a strong effect either way: **the Reflect condition is indistinguishable from the non-Reflect condition on all self-report measures** (Table 3.5). If we break out the data over time, in the first half, the Reflect condition performed worse than the non-Reflect condition, while in the second half, the Reflect condition performed better. A tempting interpretation of this data would be that as a discussion grows longer and less tractable, participants in the Reflect-enabled discussion perceived that others were extracting useful takeaways and thus had a more favorable impression of the discussion. To support this hypothesis, we would have expected a gradual increase in the Reflect conditions'

Table 3.4: Exit survey questions.

Rate how productive this discussion was (1-5, 1=low, 5=high)
Rate how much you enjoyed this discussion (1-5, 1=little, 5=great)
Rate how much you liked the other people in the discussion (1-5, 1=little, 5=great)
Rate how insightful you found other people's comments (1-5, 1=low, 5=high)
Rate how useful you think the discussion would be to amazon (1-5, 1=low, 5=high)

Table 3.5: Results from the exit survey.

	Overall		First half		Second half	
	no-Reflect	Reflect	no-Reflect	Reflect	no-Reflect	Reflect
productivity	3.7	3.7	3.9	3.6	3.6	3.9
enjoyment	3.9	3.8	4.1	3.8	3.8	3.8
liked others	3.8	3.8	4.1	3.7	3.6	3.9
others' insights	3.9	3.8	4.2	3.6	3.7	3.9
useful to Amazon	4.2	4.0	4.5	4.2	3.9	3.9

relative favor over time. But if the data is broken out in quartiles instead of halves, the data is not consistent with this explanation (e.g. the third quartile is Reflect's worse performing quartile overall, while the fourth quarter is its best). For the time being, we must reject this possibility until we replicate the study over multiple discussions. The high-level takeaway from this comparison is that it is unlikely that Reflect, on its own, will have a drastic impact on discussants' perceptions of the discussion and other participants.

Limitations and future work This is the first experiment of its kind, and much can be improved. Our next step is to summarize the main, distinct ideas from these discussions and evaluate them. Moreover, we can present them to Amazon, who will be in a position to determine which ideas are actionable. Will the Reflect-enabled discussion have led to more actionable ideas? Second, we intend to run multiple iterations of the study; right now, the number of participants (and discussions)

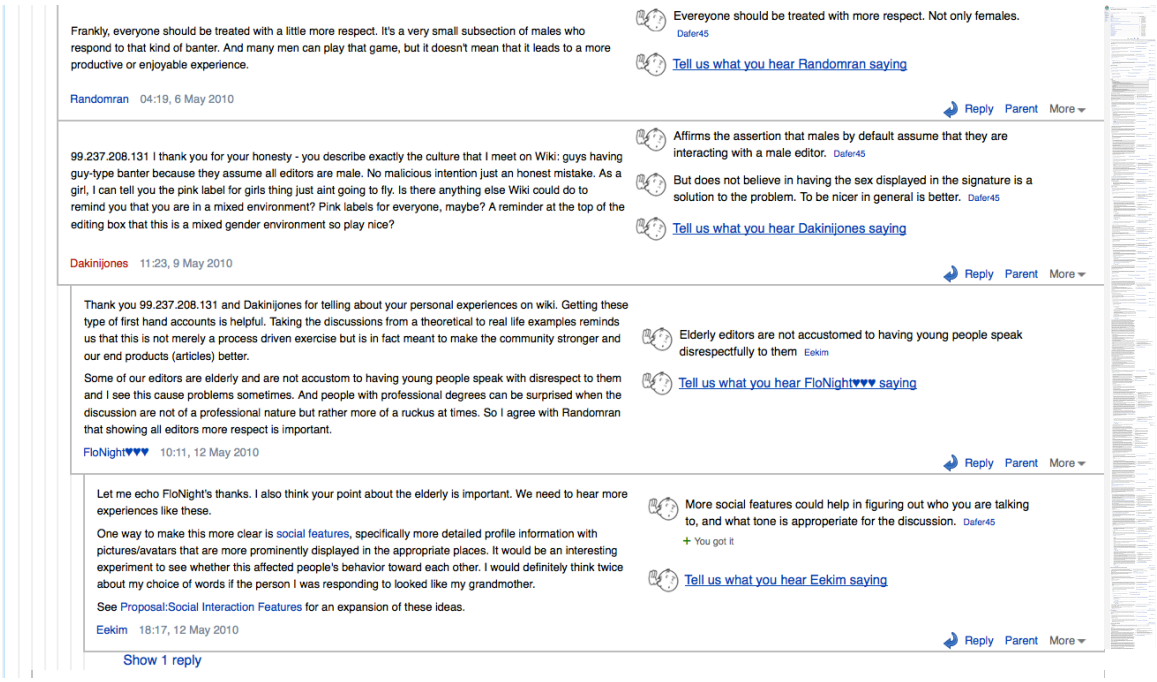


Figure 3.5: An except from one of the discussion pages for the Wikimedia Strategic Planning Initiative. As before, the full discussion is embedded on the right-hand side of the screenshot, and can be read by zooming in on the digital version.

is too low to draw conclusive results, particularly with self-report measures. Ideas workers proposed during the first run of the experiment can serve as the discussion topics of subsequent runs. Third, we wish to rectify the methodological problem with our exit survey in which participants filled out the exit survey before they experienced someone summarizing their points. This problem interferes with our ability to evaluate whether a commenter whose point is listened to will be more likely to positively perceive other participants.

Case 2: Summarizing large discussions for Wikimedia's Strategic Planning Initiative

Professional discussion facilitators often need to summarize and extract themes from long discussions in order to extract value from them and report the outcomes. This is an important but resource intensive task for effective community deliberation. Reflect may help with this summarization pro-

Table 3.6: Bullet point classification for a sample of 100 points created in the Wikimedia deployment. Adopted the same codebook and methodology reported earlier.

Code	Prevalence
Reflects back meaning of the comment	97.0%
Pure reflection that neutrally restates without elaboration	92.0%
Expands meaning by reading between the lines	4.0%
Meta observation that contextualizes the content	0.0%
Antagonistic or cheeky reflection of meaning	1.0%
Contains a reply in addition to reflecting back meaning	0.0%
Doesn't reflect meaning contained in the comment	3.0%
Flames the commenter with an ad hominem	0.0%
Off-topic and doesn't relate to the discussion	0.0%
Replies to the commenter without reflecting meaning	0.0%
Repeats simply repeats verbatim part of the comment	3.0%

cess, as it allows facilitators (or participants, depending on the conditions of the deployment) to create intermediate summaries at the comment level. These intermediate summaries can then be clustered and synthesized to create the high level executive summary.

To gain insight into this potential use case, we were fortunate enough to get involved with Wikimedia's Strategy Planning Initiative (SPI). Wikimedia, the organization behind Wikipedia and its other global sister projects, ran the SPI from mid-2009 to mid-2010 in an effort to gather input from volunteers worldwide about the strategic directions the organization should take for promoting Wikimedia projects and its respective communities. Near the end of the initiative, Eugene Kim, the lead collaboration consultant contracted to design and lead the SPI, contacted the first author about using Reflect internally to process and summarize the input they had been receiving.

During June and July 2010, five facilitators and volunteers installed a Greasemonkey implementation of Reflect, a Firefox plugin that allows the modification of any website. *This meant that Reflect restatements were only visible to the people who installed the Greasemonkey script*.

These facilitators created 282 unique summary bullets to aid their summarization task. See Figure 3.5 to see these facilitators using Reflect. Just like with Slashdot and Mechanical Turk, we coded a sample (100) of the bullet points. See Figure 3.6. Of all the deployments, these were the most straightforward bullets, with 97% of the bullets reflecting meaning (and 92% pure reflections). This is unsurprising given that the facilitators decided to use Reflect with the explicit purpose of summarization. Perhaps more interesting than the raw breakdown is the summarization workflow that facilitators reported to us. In particular, while the wiki software allowed discussants to summarize *threads* of comments, no one was doing it because, perversely, the more comments the thread contained, the more work it was to summarize, even if it was most valuable. As Eugene stated in an interview: *"We were encouraging facilitators to constantly summarize discussion threads as a way of capturing all of the key nuggets in one place and of refocusing the discussion. This was a long, laborious, manual process, and the tool we were using (the LiquidThreads extension on MediaWiki) did nothing to make this process easier. Reflect addressed this specific problem by encouraging people to summarize individual posts, which in turn made it easier to read through and summarize longer threads."* Specifically, facilitators copied and pasted bullet points into the thread level summary, and then synthesized them. This is indicative of an important future direction for Reflect: development of and explicit support for a summarization workflow integrated with Reflect bullet points.

Here is a public post that summarizes the value the SPI found in Reflect:

> We used Reflect as part of Wikimedia's open strategic planning process, an unprecedented year-long effort that drew over a thousand participants. As you can imagine, continuous synthesis was a critical component of the process, and it required a Herculean effort by some of our volunteers to summarize some of our discussions. That changed when we installed Reflect. Reflect created a lightweight way for summarizing individual posts in a participatory way. Having more people summarize this way vastly increased people's abilities to synthesize long, complex threads. I believe it had an impact on the quality of discussion as well, although we didn't have a chance to thoroughly evaluate this. I think Reflect is particularly powerful exactly because it takes one – and only one – small, but very concrete step toward augmenting a conversational

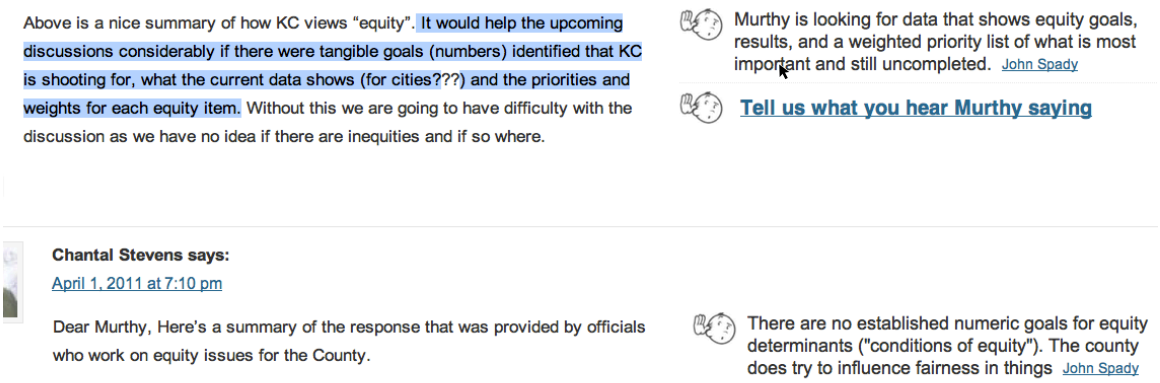


Figure 3.6: Screenshot from a CCF discussion where JS is restating points of participants in a discussion. Unfortunately, there is not a longer screenshot available due to a database problem resulting in the loss of a number of blog posts on the CCF page.

style that people are already familiar with. That approach is more likely to scale than approaches that attempt to take an all-encompassing approach to shaping conversations.

Case 3: Listening by leaders, in Countywide Community Forums

Community deliberation is often part of a decision-making process coordinated by some organization or leader. One of the major challenges these efforts face, particularly governments, is that constituents or community members feel that there is no reason to participate because they believe their voice will not be heard by those in power [92]. Reflect might help counteract this problem, at least for efforts that genuinely want to engage the community: leaders and facilitators can show that they are listening simply by restating points, without necessarily having to respond at that time with promises of specific actions. These visible acts of listening by leaders may incite others to participate as well.

To illustrate this use case, we were originally intending to integrate Reflect with the IdeaScale ideation platform and deploy it with the US Federal Communications Commission, which at the time was gathering public input regarding Net Neutrality legislation. The social media coordinators at the FCC were excited about using Reflect to show participants that they were listening, without needing to reply to each person. However, the public outreach effort fell through as the FCC director

got embroiled in a major congressional political firestorm over his initial Net Neutrality proposal. Instead, we partnered with the leader of the Countywide Community Forums (CCF), a local civic engagement effort with the mandate to stimulate discussion about issues relevant to citizens of King County, the 14th most populous U.S. county. The initiative is a public-private partnership, headed by volunteer John Spady (JS), established in 2007 with the passing of the "Easy Citizen Involvement Initiative" King County ballot measure. About once a quarter, CCF selects a new pertinent topic, and provides support for citizens to host events at private homes, libraries and elsewhere to discuss the current topic. Surveys are distributed to participants so they can record their opinions about the issue. At the end of the round, JS and his deputy coordinators synthesize these surveys into a report. These reports are circulated in the media, and JS presents them to county elected officials.

JS contacted the first author about installing Reflect. His overall interest was *"trying to get people to slow down and consider the point of view of the commenter – to honestly reflect back what a person was trying to say – before posting their new or reactive comment"*. In addition, JS wanted a way for him to help show that he is listening to participants in the initiative. For the past several of its rounds, CCF has been trying to also engage participants online through a Wordpress site, alongside their face to face efforts, but the website has struggled to gain attention (most posts receive 0-3 comments, with a few having > 10). JS therefore wants to try to cultivate community as much as possible around the comments the site receives, with Reflect as one part of the strategy (*"When I Reflect, I'm doing it with a sense that I want to engage the unseen readers. It's not so very much for the original author – except to try and let him/her know that someone is reading and responding to his/her comments."*). JS created 37 Reflect summary points, covering every comment, between August 2011 and September 2011. See Figure 3.6 for a short excerpt of one of the discussions. Overall, JS' impression is that *Reflect would be good for all types of discussions. Just having the invitation to "slow down and reflect" on what is being said is an important service.* This was our least used deployment, and unfortunately an error led to database loss, so we cannot code the bullets. However, it appeared that JS was creating bullet points roughly similar to the Wikimedia effort.

Interviews with four discussion facilitators about their use of Reflect

We conducted a structured interview via email with four expert users of Reflect: JS from CCF, and Eugene (EK), Philippe (PB) and a volunteer participant Kristofer (K) from the SPI. All of these users employed Reflect in their role as discussion facilitators. Using an email interview protocol, we asked the same four sets of questions of each participant. In the initial email, we also attached a full-page screenshot of a Reflect-enabled discussion in which they had participated so that they could better recall their experiences and therefore ground their responses. These interviews are not meant to result in generalizable, objective knowledge, but rather provide insight into how some professionals with experience using Reflect view its affordances and limitations.

The interview questions were divided into four groups. First, their experience writing Reflect bullet points, such as *"Who was the primary audience for the bullet points you wrote?"* Interviewees had the most to say about these questions because writing bullets dominated their Reflect usage. Second, their experience having a comment they wrote summarized by someone else, such as *"What was your reaction to having someone restate what they thought they heard you say?"* They were instructed to ignore these questions if no one had summarized a point they made, two of whom did so. Third, their experience reading a Reflect-enabled discussion, such as *"Did having comments with restatements on the right change how you read comments?"* Finally, we asked a set of more speculative questions about what they perceived as Reflect's affordances and limitations given their professional experience, such as *"What other kinds of discussions do you think Reflect would be good for? Why? What kinds of discussions would Reflect NOT be good for? Why?"*

Because the interviews were fairly short and we only had a few participants, we have included the entire interview transcripts (with their permission), in Appendix A. In addition, the first author has extracted a few of the salient themes that emerged from the interviews:

There were diverse motivations for writing bullets & distinct perceptions about who the intended audience was. All four interviewees created bullets to some extent for other readers. For JS, it was to model reflective behavior to other readers; for EK and K it was to help newcomers to the discussion catch up, or to help other participants summarize. However, PB primarily used it as a note-taking device for himself, particularly for comments left by non-native English speakers; he didn't bother writing reflect statements for comments he found easy to understand, unless *"a point*

was particularly good and I didn't want to lose it". Moreover, he was also *"less concerned with the author of the post, though I did several times clarify misunderstandings"*. This was due to the closed nature of the Wikimedia deployment, as he later suggested that one of Reflect's major future opportunities would be helping confirm the points non-native English speakers were making in a *"non-threatening and non-judgmental way"*.

The orientation all participants took toward writing a bullet was to distill meaning from long passages into succinct statements. In response to a question about what makes a good Reflect point and how they approached writing a bullet point, JS wrote *"Short and sweet! Trying to write the least I can to capture the essence of a portion of the original document"*; PB writes *"It's almost like writing a tweet: short, to the point, and provided the necessary info to recreate my thought process later"*; EK states *"My goal was always to highlight the essence of the post, no more, no less. I'd try to do it in a way that captured the tone of the original summary"*. K said much the same, but added that *"what was good with Reflect was that it was possible to strip away all the 'proofs' of the statement and express the idea in a single sentence, but still having the text that the bullet was derived from highlighted when a user hovered over it and thus making it possible for anyone critical of the statement to investigate the reasoning more thoroughly."* In other words, Reflect provided an opportunity to express meaning in the absence of justification.

Reflect caused all four participants to consider more deeply the meaning the commenter intended to convey. At least with these four users, Reflect provided a means and a reminder to use and hone their active listening skills. Each of the participants wrote that Reflect led them to *"pause and really listen to what the other person is trying to say"* (JS). PB states that *"We tend to jump instinctively on comment threads and Reflect forced me to stop and consider the underlying meaning of what the commenter said, as opposed to just what I initially read."* Similarly, EK writes *"Occasionally, I would read a sentence that I didn't completely understand, and I would generally just gloss over those. You can't do that when you're looking to summarize."* K adds that the highlight functionality was important to this outcome, as *"sometimes I wrote a bullet point that I couldn't highlight, and that often meant that the bullet point wasn't all too accurately formulated, and I had to reconsider if I actually had understood what the reader really meant yet."* This experience of summarizing other comments (and having others summarize their own) was powerful enough for K and PB that they started to alter how they wrote comments so that their comments were clearer (*"I*

found myself editing comments afterwards for style (removing subordinate clauses, difficult punctuation, etc)"). Overall, these responses are encouraging because they suggest that Reflect may act to counteract our tendency toward knee-jerk reactions.

There are other interesting themes, and divergent perceptions, that can be found in the interview data. Readers are encouraged to examine for themselves the full interview data (Appendix A).

3.5 Discussion and conclusion

We do not consider Reflect to be an endpoint. Rather, it is a building block for social infrastructure that enables more effective web-based public discussion and civic engagement. Specifically, we intend to extend Reflect to support crowd-sourced summarization that yields a Wikipedia-like summary capturing the most important takeaways from the discussion. In some cases, this summary might focus on actionable ideas. In others, it might capture personal experiences from community members regarding a common problem. It could be gradually written over time by many readers doing small tasks that are automatically combined (e.g. incorporating a set of restatements into the summary, verifying that a restated point is reflected in the summary, or improving the flow of the summary). Finding ways of identifying tangible outcomes from discussions will make it more appealing to incorporate community deliberation into decision-making.

This chapter has made several contributions to move us closer to this vision, describing a theoretically informed framing of the design space of listening interfaces and contributing Reflect, which enables people to demonstrate evidence of listening through restatement. Our empirical work suggests that Reflect, and listening interfaces more broadly, are a potentially useful design direction for improving public discourse on the web. In the Slashdot deployment, we showed that people will use Reflect often, and in a way that often reflects meaning (as opposed to replying to or antagonizing commenters). Our three deployments for community deliberation were more limited, and the results more speculative, but help illustrate different use cases of Reflect.

Our results, however, are limited because they do not rigorously answer the hypothesized outcomes regarding listening interfaces laid out earlier in the chapter. One major challenge for future work undertaking more rigorous analysis is that the topic at hand and the people involved in the discussions can greatly impact the interpersonal dynamics of the discussion, even in the same context.

Our Mechanical Turk deployment/experiment is a sketch of a methodology that may be extended to help test the outcomes, overcoming the challenges of ecological validity of a lab-based protocol. But much more attention needs to be focused on developing the methodology and carrying it out.

A second challenge for rigorous empirical analysis is that there are innumerable contexts in which discussions take place from news sites to classrooms to board rooms. We have presented Reflect as a general purpose discussion interface functionality in the hope that designers working even in domains we have not considered might be inspired to apply some of the principles of listening. Overall, we suspect that Reflect has the most potential for shifting patterns of dialogue in situations where (1) discussants share common social bonds such as identifying with a shared community (e.g. Slashdot), (2) discussants share a common a goal (e.g. citizens communicating with government about how to address a specific issue), or (3) decision makers employ Reflect as a mechanism by which to explicitly listen to constituents. Conversely, we do not expect that Reflect alone would do much to improve many discussion sites of well-known news sites like the Washington Post or Seattle Times, or when the topic of the discussion is abstractly ideological (e.g. a debate about abortion), although it may be part of a larger solution. We also asked the participants in our interview study about their perception of the discussions in which Reflect might improve its quality, and they broadly suggested that they believed any conversation with long posts would benefit, with K suggesting that it would be particularly useful *"In discussions where it is difficult to formulate yourself well because the concepts are difficult, I think Reflect can offer an insight into if you succeeded in getting your point across"*. PB, however, raised the concern that *"Anything that required a complex thought would tend to get distilled into a 'comic' summation of what it actually is, I think"*. This is an important possible limitation to investigate further.

We encourage other researchers to further this vision theoretically, methodologically, and through new design, evolving the web into a platform that incentivizes not only free speech and open participation, but also attentive, constructive, and inclusive dialogue through new forms of active listening.

Chapter 4

LESSONS LEARNED FOR DESIGNING FOR REFLECTION

4.1 Introduction

In this dissertation, I have thus far framed a design space for supporting more effective public discussions, highlighting two important challenges: (1) a dearth of interface support for empathetic listening, consideration of tradeoffs and introspection, which undercuts our ability to find common ground, and (2) few interface mechanisms beyond social filtering for identifying valuable takeaways and maintaining the interpretability of a discussion so more people can participate effectively. In particular, I have identified an opportunity for addressing both of these seemingly disparate challenges through a common strategy whereby the use of reflective mechanisms might not only improve interpersonal dynamics but the output may also be used to help scale discussions. Reflective mechanisms can thus play a dual role, and can be described as acts of facilitation. To inquire further, I designed and built two novel systems premised on this design strategy, demonstrating through field deployments that the reflective mechanisms do in fact have traction for wide public use, showing that crowdsourcing facilitation is a fruitful direction to pursue.

While Reflect and ConsiderIt share this common design strategy, each of them followed different trajectories through the design space. Despite these different trajectories, I converged upon a couple common design themes without explicit intent. In this chapter, I discuss two of these themes as lessons that I have learned about supporting reflective public thought. By conveying some of these lessons, I hope others will be able to better approach this design space, because it is not easy to (1) nudge people to reflect and (2) do it through simple designs.

The first lesson is that articulating utterances as length-limited, semantically constrained "points", such as pros and cons or Reflect summaries, confers a number of benefits related to lowering the barrier of reflective contribution, including more people into the reflective activity, and ultimately providing leverage for surfacing common ground. The second lesson is that exploiting column-based layouts can structurally make reflective activities understandable, actionable and enforceable. In presenting these two lessons, I discuss a number of relevant design dimensions and where prior systems can be classified along these dimensions.

4.2 Lesson 1: Articulating utterances as length-limited, semantically constrained Points provides leverage for lowering barriers to reflecting, involving more people, and surfacing common ground

In discursive spaces, discussants must be given the means to express their thoughts, whether they are first-order comments or reflections upon others' comments. The constraints the interface imposes on these utterances bounds the communicative affordances of the space. In other words, the means by which the interface articulates the work of communicating shapes the discursive possibilities.

In this section, I first discuss a few design dimensions related to articulating utterances: constraining their semantic content, constraining their length, and limiting how they can be updated over time. In the course of discussing each dimension, I characterize previous approaches and reflect upon the choices I have made in my design inquiries. I conclude the section by discussing the lessons I have drawn from using a common strategy for articulating utterances with both Reflect and ConsiderIt.[1]

4.2.1 Domain of utterance

Discussion interfaces can vary in terms of their explicit demands for utterances to conform to particular types.

Open Many web-based discussion interfaces do not impose any restrictions whatsoever on what is to be said. Blog comments, for example, generally do not dictate anything about the substance of the comment, beyond the implicit mandate that the comment relates to the parent post.

Semantically constrained Discussion interfaces can, however, explicitly specify a semantic constraint on an utterance, and potentially provide mechanisms by which to enforce it.

Some systems draw on theoretically situated semantic structures to constrain utterances. For example, the argumentation mapping line of research for collective sensemaking where people are forced to break up their narratives into fine chunks, each of which the user must classify according to an argumentation schema based on theory (e.g. gIBIS, Cohere, and hundreds of others, as re-

[1]Note that I only consider textual utterances, although there are interesting possibilities to consider in the future for video and audio utterances.

viewed in Section 1.1.3). These theoretical structures sometimes cause problems for users who are unfamiliar with the constraints. Other systems use more familiar semantic constraints, such as the Q/A format, ConsiderIt's Pro/Con format, or Idea platforms.

By formalizing utterances through these semantic constraints, designers gain representational and computational leverage. For example, if participants are asked to enter a "Pro" or "Con" that can be included in other people's pro/con lists, ConsiderIt can computationally identify the most salient pros and cons across different user groups (PointRank), or, representationally, show other pros and cons in the margins of the pro/con list. Similarly, with a Q/A site, by restricting replies to questions as "Answers", the ability to identify and privilege the "best" answer is afforded, as well as enabling reputation systems that recognize the best answerers. In short, semantic constraints provide design leverage.

However, as Shipman [88] eloquently describes, introducing these formalities can have harmful consequences. Human communication is nuanced, and formalisms, no matter how theoretically rigorous, are always constraints on open narrative. There is a tradeoff between expressiveness and design leverage. This is most clear with the argumentation mapping line of research. Though based on multiple lines of rigorous argumentation theory, few people (1) understand these abstract linguistic semantics or (2) understand how to apply them when breaking down their own utterances, or other users' utterances, according to the linguistic semantics. Despite hundreds of systems designed in the argumentation visualization line of research, there have been no field deployments suggesting that anyone except expert users would ever use them. ConsiderIt also imposes an argumentation model (pros/cons), but because nearly everyone is familiar with pro/con argumentation, it appears that it was successfully appropriated. As P3 in our user study said, *"What I really like is that its not confusing, its nice and presentable, that's how I like everything, just nice and clean and easy cut, not all these crazy things"* , while P7 said *"the way it's set up is very clear, concise. Everything's organized"* .

But even with more familiar semantic constraints, problems arise as users struggle to align their utterances with the semantic constraints. Despite the wide understandability and relatively high degree of flexibility of the pro/con format, certain valuable classes of utterances were not accommodated, such as questions about the ballot measures or neutral informational content. Moreover, the pro/con list format does not enable people to connect different points to each other, or establish

hierarchical relationships, as the more semantically flexible argumentation visualization systems do. For example, P3 in our user study wanted to essentially add a counter-point to a point, but there is no facility for linking points in a traditional pro/con list (*"If anything I would put it as an attack on con, but its not really a pro...its kind of a counter on what other people have been saying"*).

There was also evidence that users struggled at times with semantics of Reflect bullet points. For example, some users apparently found Reflect bullet points to provide an intriguing means by which to provide a short response to a particular part of a comment. A community moderation mechanism worked fairly well to identify and deactivate these bullet points, but the very presence of the replies suggests that users sometimes wanted to appropriate the functionality in ways that went against the "restatement" formalism. Similar issues arise for Idea platforms and Q/A systems. Consider Obama's Dialogue on Open Government, where users often submitted "ideas" which were not actionable, but rather abstract principles like "Government should be more transparent".

The tension between expressibility and formality is difficult to navigate. However, by using formalisms that are part of commonly known formats many of the problems arising from formalisms can be avoided because at least users understand the baseline format. I have tended to trade off semantic flexibility for a tool that people can and are willing to use. I recommend observing that statements users make through the interface and consider extending the interface to support those communicative uses which are restricted if enough users appear to be desiring that means of communication.

4.2.2 Length of utterance

People do not have the time or patience to parse long discussions, so it is important to maximize the use of their limited attention. Utterances can be constrained in their length such that utterances take up less room so that a more inclusionary space is created.

Unconstrained length Most web-based discussion interfaces do not constrain the length of an utterance. However, a commenting space will then quickly succumb to massive amounts of text, constraining the ability of participants to make sense of higher level patterns in what people are saying, or engage with more than a few comments. A common pattern for dealing with this information overload is social voting, where users vote on utterances and the highly rated utterances are privi-

leged in the interface, such as showing up higher in a linear list (e.g. answers in Stack Overflow) or even being the only ones displayed (as with Slashdot's filtering approach). [2]

Restricted length Another approach is to limit the length of utterances, such as by restricting utterances to a certain number of words or characters. While this strategy has not been extensively explored, Twitter's discursive ecosystem demonstrates its potential. I have used this approach with both ConsiderIt and Reflect, and I have found that it provides great design leverage.

First, the length limitation forces users to think hard about what they actually want to articulate. See for example the interview quotes in Section 3.4.2 about how Reflect focused users on distilling the essence of a comment (e.g. *"My goal was always to highlight the essence of the post, no more, no less."*. Second, limiting the length of utterances saves real estate. With a 140 character limit, more statements by more people can be shown at once in the same visual frame. Without character limits, a single user can take up a great deal of room with a single rambling comment, which can undermine the integrity of a discussion by driving away potential commenters and tainting later comments [93]. The length limitation is thus possibly a reminder to be polite to others: everyone is busy, so you should try to convey your point as succinctly as possible, and make room for others. Third, limiting the length conveys to users that they should be articulating singular points at once, instead of weaving together a longer narrative. For the particular types of reflective utterances I was trying to support, this was critical.

Limiting length also allows designers to make assumptions about the visual size of the utterance in the interface so that more radical discursive layouts can be employed. This is readily apparent in ConsiderIt. When users craft their own pro/con list, ConsiderIt is able to cram *eight* pros and cons in the *margins* of the list, all in the same visual frame. This is essential for enabling the primary UX of personal deliberation augmented by the thoughts of other people. Moreover, without the length limitation, the point inclusion functionality would have been difficult (or even impossible) to effectively design. As P1 in our user study suggested, *"I really like bulleted points that are short so*

[2] Another pattern for dealing with longer comments is to arbitrarily limit the amount of text shown per comment, with the option of clicking "read more". As I describe later, I prefer to ask users to write a "nutshell" version of their comment, which is shown alongside everyone else's summary, and then users can optionally enter a more detailed comment. The reason I prefer this option is because the user is forced to reflect on the point they are trying to make, and because users browsing the text will be able to read a fully formed thought, rather than just the first few characters.

it's nice to get a brief idea of what people are saying and I like the simplicity of how this is set up."

Limiting the length of utterances, however, has a clear drawback: it can prevent people from articulating complex narratives that weave together multiple thoughts, thus limiting the possible nuance conveyed in an utterance. This is evident in Twitter, where in-depth conversations often seem stunted because of the short messages and users' desire to not spam all of their followers in order to overly engage with one or a few users. In ConsiderIt and Reflect, however, I created a pattern to mitigate this problem: enable a succinct, length-limited summary alongside a detailed response.

In ConsiderIt, users were asked to write a short "nutshell" description of a pro or con point they were authoring and, optionally, add a longer detailed description that elaborated upon the nutshell. The nutshell description was shown to other users who were browsing, and anyone could click on a "read more" link that expanded out the detailed description, enabling a summary+drilldown functionality for managing screen real estate while still enabling more nuanced articulation of the point. In the 2010 LVG deployment, points averaged 104.4 characters (min=21, 20.6% $\leq$ 135 characters) and only 45.5% contained a long description. Authors did not seem to be overly constrained, and used the point detail functionality when they found it necessary. As P1 states, *"there's an option for a condensed version and then, um, like a bit more detailed level in case you do need to go into more detail about you're point, so I think that's nice."* While most users understood this, it was clear that some users had difficulty in understanding the distinction between a nutshell summary and a long description because a number of users left the summary blank or simply wrote the first part of their point in the summary field and continued into the details field.

In Reflect, the pattern manifested in a more interesting way. Responsibility for creating the high-level summary of a detailed comment did not rest with the commenter. Rather, Reflect asked "listeners" to write the "nutshell" distillation of the longer, more detailed utterance of the commenter. If desired, readers of the summaries could optionally drill-down into the details by examining the highlighted text of the source comment. Recall K's quote in Section 3.4.2 of the Reflect chapter about how this aspect of the Reflect functionality was greatly appreciated.

In the current design of Reflect, commenters cannot restate their own points. However, it would be interesting to create a comment platform that worked more like ConsiderIt's nutshell+detail pattern: commenters summarize a series of distinct points they wish to make, and can optionally write

a more nuanced full comment. Readers browse the summaries, optionally drilling deeper to hear more if interested. This would undercut the listening and grounding functionality of having others write the summary, but it would perhaps lead to more coherent overall discussions than the status quo.

4.2.3 Authorship of utterance

The management of utterances can also vary based on who is authorized to modify them. Although I employed the common approach of just allowing the original author to modify a point, there was extensive deliberation about the choice. Moreover, I have identified some potentially high-value variants to pursue in the future that I will briefly cover here. These alternative models of authorship may provide leverage on synthesizing the thoughts of multiple people in a single utterance, which can also help address the problem of information overload.

Closed group The most common and straightforward approach is to restrict modification to the original author, or a closed group of authors, or even prevent modification entirely. In ConsiderIt and Reflect, I have not strayed from this authorship model. But there were some thorny tradeoffs in making this decision.

In Reflect, there was an interesting tension in authorship between the commenter and the reflector. One might argue that a commenter should be treated as a co-author, or even the primary "owner", of a Reflect bullet point, given that a bullet point is intended to reflect back the meaning of the commenter's point, and is given fairly prominent screen real estate. However, I did not feel that *interpretations* of the comment should fully be under the control of the commenter, because (1) misunderstandings might actually be useful for other users to see and (2) because reflectors could call out implicit or inferred meanings underlying the comment that the commenter purposely left implied. Even though this use of Reflect could sometimes be antagonistic, it can also serve to help surface deeper underlying value tensions in the discursive community. Were commenters to be seeded control of Reflect bullet points, they might use it to suppress dissent or undesired observations. I believe that the Slashdot deployment vindicated this choice: the (often cheeky) bullets that read between the lines or were meta observations often provided the most insightful and useful interpretations of the comments (at least from the coder's perspective). Because they sometimes

cast the commenter in a negative light, they probably would often have been edited or removed by the commenter. I mitigated the tension between authorship and reflective utterances by enabling commenters to clarify their points, which are also shown prominently to other users. This seemed to work well, but there are likely other strategies for addressing these issues under different circumstances.

In the 2010 LVG deployment of ConsiderIt, point authors could not edit their own points. This was due to a general tension that arises anytime an utterance might be reflected upon: what happens if the author radically modifies their utterance after a number of people have already voted on/included/restated the original utterance? This is, in other words, a potential line of attack for malicious parties, where the attack is to be the first to state a salient point, let many people recognize its importance, and then change it to something else, perhaps offensive content or a way of trolling its former supporters. I call this the *trojan point* attack. The same issue arises in Reflect: what happens if a commenter updates his or her comment after someone reflects on it? What happens if a reflector updates their bullet point after the commenter responds? In Reflect, I simply did not let users update their bullet points after the commenter responded, but this just avoided the problem and was unsatisfactory. In general, this is a thorny issue that many designs need to be able to handle if dynamic, evolving utterances are to be supported alongside supporting reflections upon those very same utterances. Devising general approaches to this problem is a key direction for future research.

Gate-kept An utterance might be controlled by the original author (or collection of trusted authors), but others can recommend changes that the original author can accept, similar to a patch submitted to an open source code repository.

An interesting extension of ConsiderIt based on this model would be to enable users to restate a pro or con point when they include it into their lists, popping up a dialog box after the drag-and-drop. If the includer decides to restate the point, the original author might be notified, and, if they believe that it is a superior articulation of the point they were trying to make, they could then designate it as the primary instantiation of the point, which then gets shown to other users. This is, in essence, a completely different method by which to enable the basic Reflect interaction model. It would have a number of benefits:

1. Creates the possibility for the evolution of a pro or con via community wisdom without forcing the original author to give up expressive control. Right now, an author might raise an important point, but its tone may be off, or it confuses some aspect of the issue, or is not succinct. This method could correct these problems over time.

2. This functionality would engage both communitarian communicators, who would like to see the points representing their own side presented in maximally persuasive ways, as well as deliberators, who may be interested in showing themselves and others that they understand a point by restating it. Aside from making it more likely that someone rephrasing a point would experience an opinion change (as described in Section 2.5), the fact that a group of people is actively evolving the point might lead to an increase in community efficacy.

3. Provides a purpose for notifying the point author about someone listening and reflecting upon a point they had authored. To date, ConsiderIt has not notified point authors of an inclusion of their point, or even provided any means to learn when or how often their points have been included in others' lists, because including a point does not seem to warrant an email notification (and displaying numbers related to inclusions added too much visual clutter). This undercuts the potential benefits of listening as described in the Reflect chapter. But with the additional functionality around rephrasing, a suitable excuse for notifying users about acts of listening would be created.

Open An utterance might be collaboratively edited. This was the original idea for a Reflect summary (see Figure 4.1; at that time, it was called "Extinguisher" and it was intended as a means for a community to rewrite the entirety of a comment in order to "civilize" it). With respect to ConsiderIt, a pertinent comparison is Ballotpedia, a collaboratively authored wiki whose main purpose is to collect information and document pro/con arguments being made about ballot measures.

In general, the collaborative approach has the advantages that work is less likely to be duplicated and, if executed properly, results in a synthesized artifact that may often be more readily made sense of than a standard web forum. However, there are some downsides. First, this collaborative approach frames the activity more as documentation or mapping, rather than an expression of one's

own thoughts. This limits the opportunities to express one's own opinion, as it implies that if the argument has already been articulated, then it should not be expressed again. Of course, users can edit what others have said before, but this functionality introduces its own social costs and benefits for negotiating how it is said. Finally, confusing user experiences are the norm for collaborative editing systems that support strangers mapping out contentious issues; in these circumstances, complex and unfamiliar mechanisms need to be supported (e.g. edit histories and reverts in wikis).

But there are some potentially high-value applications of an open approach. I review a number of possibilities for a collaborative editing approach that builds from Reflect and ConsiderIt in Section 5.1 of the concluding chapter. Moreover, at the end of this section, I illustrate another point in this design space that employs open, collaborative-editing in the public deliberation space: enabling dynamic *cohorts* to author position statements.

4.2.4 The virtues of articulating utterances as length-limited, semantically constrained Points

I have thus far described three dimensions related to articulating utterances: their semantic type, their length, and who can modify them, as well as some aspects of how I explored those dimensions through ConsiderIt and Reflect. In this section, I reflect upon a high-level lesson that I have learned with respect to all three of these dimensions.

In Reflect and ConsiderIt, utterances were all (1) semantically constrained (pro/con, or restatement), (2) length limited, with drill-down capability, and (3) closed authorship. The underlying design theme became "points": users created Reflect summary bullet points or ConsiderIt pro/con points. The systems could then enable interactions around those succinct, semantically constrained utterances. This provided great design leverage.

To illustrate the design leverage with respect to Reflect, consider an earlier design: instead of a list of succinct bullet points, I had originally designed a wiki-based summary, i.e. (1) a semantically constrained (summary), (2) with no length restriction, (3) that is openly editable by many. An early mock is shown in Figure 4.1. However, as described in the design rationale section of the Reflect chapter (3.3.2), I encountered resistance to this design during early paper prototyping, in particular, people felt that with a wiki, they must commit to summarizing the whole comment (many of which can be quite long). When I finally had the idea for a bulleted list of individually authored summary

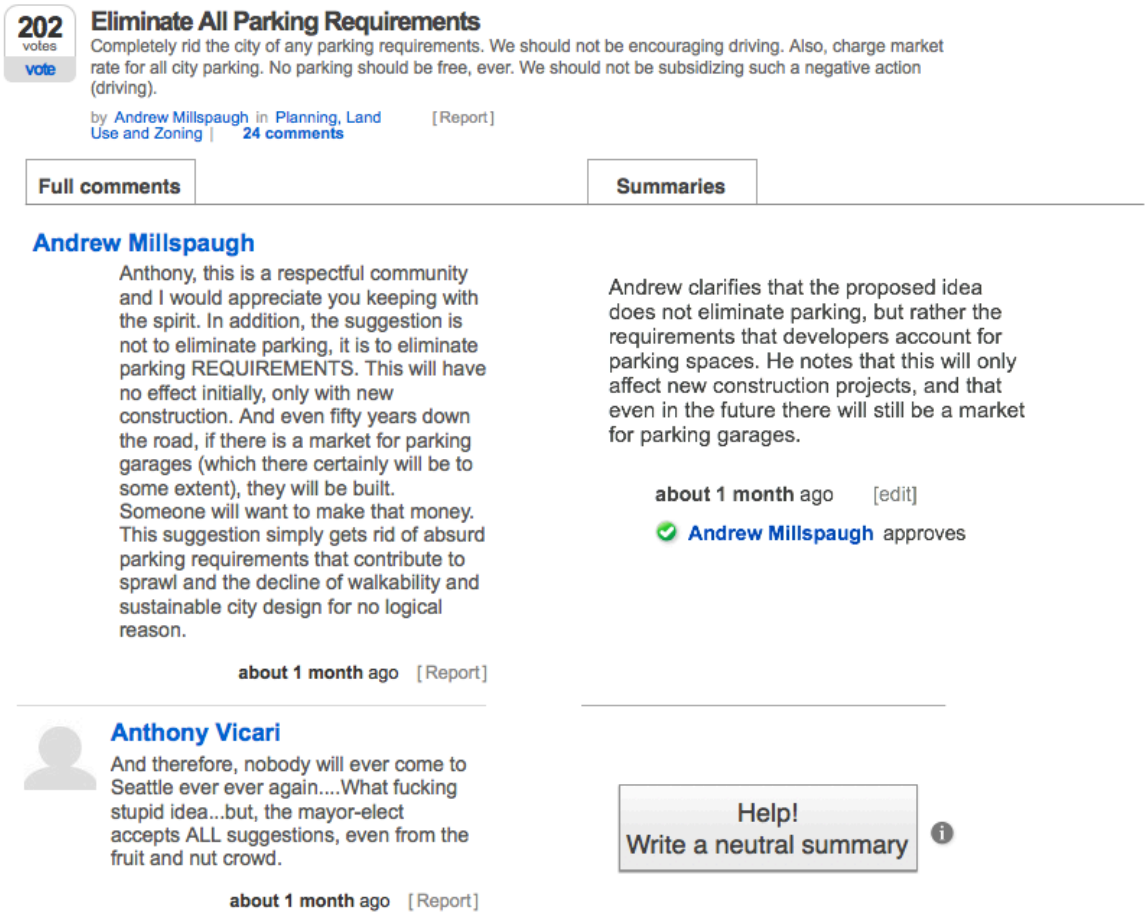


Figure 4.1: An early prototype of Reflect that used a single wiki as the comment summary.

points and started prototyping, I discovered a range of affordances: (1) we could more naturally include more people in the listening and grounding activity, (2) enhanced support for clarifying singular points, (3) facilitating multiple interpretations of the same text, even at different levels of meaning, (4) easier for commenters to respond to the summary and clarify their points, (5) easier to skim the summaries and connect them to the comments, (6) the ability to identify similar points being made in other comments, and (7) the ability for listeners to highlight the points they found most salient in what the commenter said.

More abstractly, and including some of the benefits discussed earlier, emphasizing succinct, semantically constrained points allows for: (1) saving screen real estate, (2) representational leverage

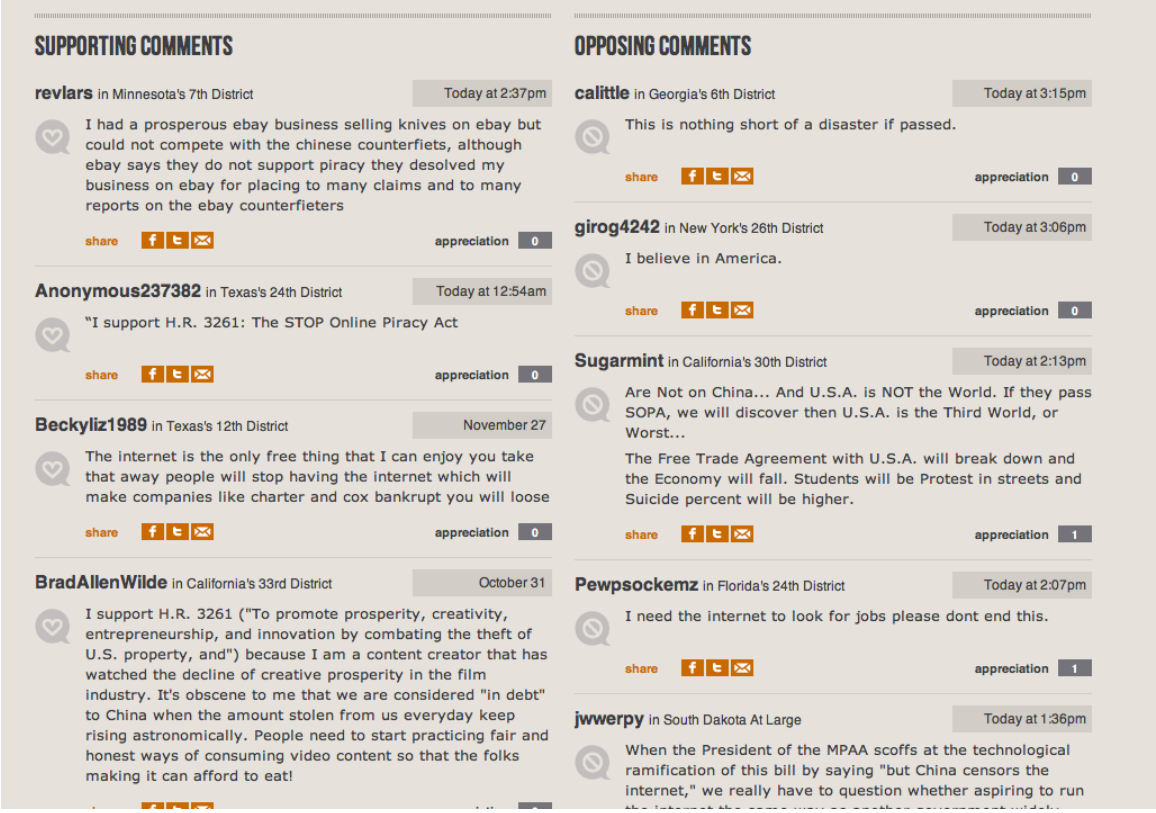


Figure 4.2: PopVox (`http://www.popvox.com`), the most similar system to ConsiderIt. Side note: observe the "appreciate" framing of the "like" button, a much cleaner articulation of listening functionality than the "like" button (which overloads judgment and listening), although the need to click the heart icon detracts from this conceptual distinction.

for radical layouts, (3) forcing people to reflect upon what they actually want to say to fit length constraints, (4) lower barrier to entry, (5) ability to re-focus discussions around singular points to maintain focus, and (6) ability to articulate the same point from multiple angles or levels of meaning.

I exploited this knowledge when designing ConsiderIt. To illustrate the power of the approach, compare ConsiderIt to its most similar project, PopVox (Figure 4.2). PopVox enables deliberation about congressional bills, but users are asked to articulate support/oppose positions of arbitrary length, instead of succinct pro/con points that other users can leverage to express their own concerns. ConsiderIt's primary interaction, personal deliberation with optional drawing upon of other

people's thoughts, is simply not possible in a system like Popvox that articulates utterances at the statement level. Moreover, because the pro/con points that users are making are mixed together across the position statements, there is no way to identify the salient concerns of the public that has formed around the PopVox page. Perhaps most importantly, without point-based articulations of concerns, it is extremely difficult to identify potential common ground. In ConsiderIt, common ground can not only be identified, but it can be brought to the fore (via PointRank's *diverse appeal* term). Overall, the experience of PopVox feels far shallower than ConsiderIt, all a consequence of the design leverage gained through a point-based style of articulating utterances and the reflective inclusion functionality. [3]

4.2.5 Openly edited position statements: illustrating a different point in the design space

To conclude this section, I sketch another point in the design space which is quite different in its approach to articulating utterances as points: enabling dynamic *cohorts* to author position statements. This is actually where I started conceptually back in 2008, when I had the insight that we could address problems of scale by reconfiguring the division of communicative labor such that those who wish to communicate their concerns to decision makers would have the responsibility to unite together and express their collective thoughts as a group. What follows is an extracted description of this idea taken from our NSF SoCS grant. By including the excerpt without further editing, I can convey here a snapshot of the underlying thinking about the problem of civic communication at an earlier point, as well as illustrate how some of the dimensions I have described in this chapter were already at play. Note, for example, an early articulation of the "trojan point" issue raised earlier. Ultimately, it was through my struggles to design an understandable version of this approach that I eventually arrived at Reflect and ConsiderIt. As will be apparent through the discussion of future work in the following chapter, Reflect and Consider have provided some leverage that may enable us to return to the original motivations described here, and actually design a usable version of co-authored position statements. Without further ado, the excerpt:

---

[3] Note that, in PopVox's defense, it does afford high-level position statements, something that earlier versions of ConsiderIt does not support, which is one of its limitations. Recently, I have extended ConsiderIt to enable these summative position statements that so that users can explain their support/oppose stance with respect to the pros and cons they included, in the context of a new system for academic peer review.

Position statements in current political discourse are often either strictly limited in number (such as the pro and con statements on the official voter's guide for referenda and initiatives), or in terms of individual contributions, such as emails and letters. We propose a more open system for collaboratively authoring position statements that are created, written, and endorsed by citizens visiting the engagement site, using wiki technology. It can be seen as an open, competitive marketplace of co-authored petitions that all bear upon the question being posed (e.g. what should City Council's priorities be for 2010? What do you think about the alternatives for replacing or removing an aging waterfront freeway?). Individuals can find and band together with other individuals and organizations to author position statements, critique the positions of competing groups, and endorse backed alternatives. These position statements are then entered as part of the public comments process. Some of them, by a process established ahead of time, might even be given privileged status by inclusion in official voters guides.

Each position statement is coupled with a discussion forum where its content can be negotiated. This model is motivated by Wikipedia, where the usually high quality article content often belies the raw content on the discussion page. In the same way, synthesis of raw opinions is likely more suitable for circulation and consideration by a wider public, because of its possible succinctness and persuasive qualities. However, unlike Wikipedia, the documents being authored are inherently advocacy-oriented, rather than striving toward a neutral point of view, which raises new issues for resolving issues of content and authorship [12].

This design empowers (and depends on) participants to self-organize. And instead of putting the synthetic burden on readers of opinions, advocates are given the ability to present their best arguments and mobilize support. We hypothesize that those seeking to attract supporters and aggregate influence for their preferred policy outcome will be motivated to make arguments more comprehensive and accessible than those found in solely combative discursive environments because of competition for endorsements. This helps to align individual motivations with the desired tasks, one of the central principles for designing groupware [42]. Compared to state-of-the-art engagement systems such as IdeaScale, we hypothesize that this technique will yield positions that are (1)

more consistent and well presented, (2) fewer, (3) demonstrative of a position's support, and (4) more persuasive.

This system opens a variety of ways for individuals to contribute to expressions of political sentiment in ways that reflect their interests and creative talents. Those who feel most qualified and motivated to create an artistic statement about an issue can work alongside citizens more interested in writing and argumentation. Others might be involved in doing copyproofing work or recruiting others to help out. This ability to breakdown tasks and reintegrate them is seen as integral for successful large-scale collaborative efforts [8]. We hypothesize that these different means of participating will help to engage actualizing citizens.

There are a number of research challenges here. First, many participants will want to have control over their expressed positions. A number of policies are possible, such as open editing (anyone can edit), closed editing (editing privilege to trusted editors, like many open source projects), and gatekept (approval required for changes). If authorial control is introduced, it will be important to have a counterbalancing freedom. All text should be free to be copied and modified. This "right to fork" is a core feature of open source projects [100]. It assures that if contributors to a position find that they have irreconcilable opinions or the initial author is found to have a hidden agenda, they are able to just start another position based off of the original. But there is a fine line between duplicated position statements that spam the site and legitimately forked positions. A design question is how to incentivise participants to fork only as a last resort.

Second, methods for demonstrating support for positions or actions might include simple voting mechanisms, submitting a short message in support, and/or becoming a co-author of the position. But endorsements are complicated because a position statement can change. How should endorsers of a changed position statement be notified, if at all? Should there be a call for official endorsements when a "final" version has been established? Moreover, the voting mechanism is vulnerable to being gamed. How can virtual sig- natures be invested with their proper authority? Identities can be verified through rigorous methods, but providing information is a privacy concern to some, and

a nuisance to others.

Third, advocating for an outcome does not depend on everyone agreeing on the rationale. Yet providing coherent rationale for a decision is important for persuading others of the salience of the issue and the integrity of the advocated action. Therefore, to aggregate influence, individuals may want to pull multiple position statements together under one banner to form a coalition that advocates for an action to be taken. Coalitions allow momentum to build behind some actions, an important outcome of an engagement process. This dual level structure creates a unique intellectual challenge for supporting group work.

4.3 Lesson 2: column-based layouts can structurally make reflective activities understandable, actionable and enforceable

Discussions are comprised of many utterances, exchanged between multiple participants. In this section, I examine some ways in which discussion interfaces can model the relationships amongst these utterances and visualize them. I conclude with a lesson that I have learned about the virtues of column-based layouts for making reflective interfaces understandable, actionable and enforceable.

4.3.1 Modeling connections

The most common method of organizing utterances is as a hierarchical tree, with successive levels indicating a response to the parent. Most discussion systems are modeled in this fashion: traditional comment boards, Q/A systems, and argumentation systems. ConsiderIt and Reflect are also modeled as trees, with, e.g., Reflect bullet points rooted to the comment, and responses rooted to bullet points.

Another approach is to entirely discard any notion of connecting the utterances, and just present a space of first-order utterances. This is the strategy followed by OpinionSpace [32]. OpinionSpace models the discourse space based on the *author's* answers to questions about their values.

4.3.2 Visualizing utterances

There is far more variance in how these utterances can be visualized than in how they are modeled. I briefly describe some of the ways in which prior work has visualized utterances, before delving into the primary lesson I have learned with respect to visualizing discussion spaces.

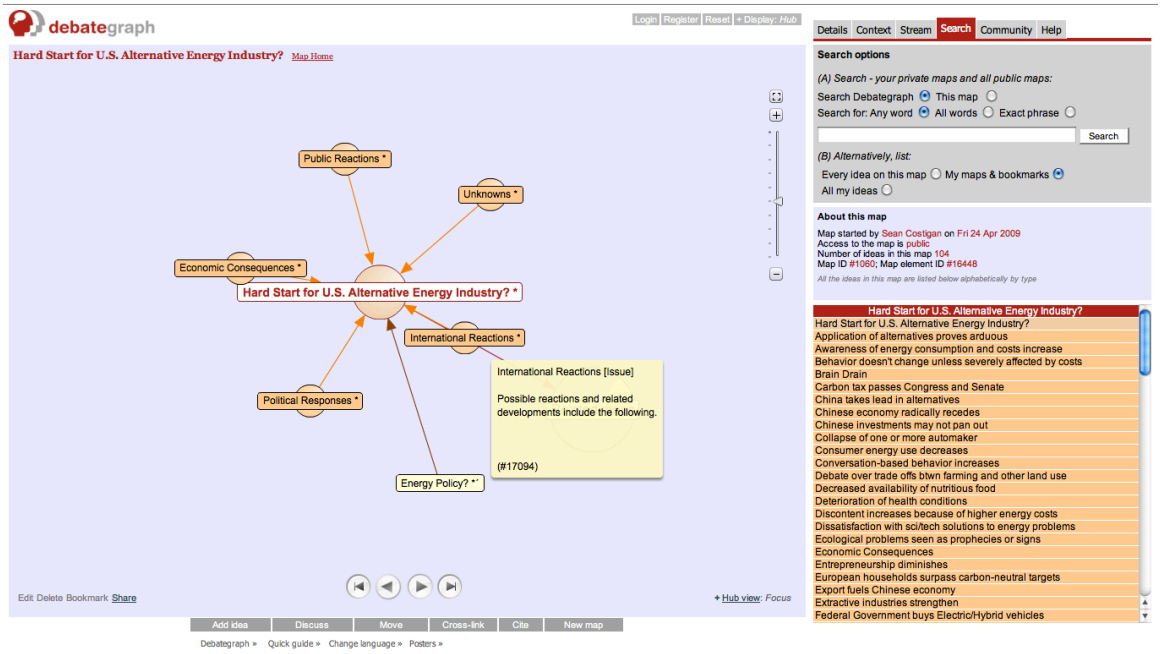


Figure 4.3: An example graph-based layout of a discursive space, from one of the most popular examples of this technique, `http://debategraph.org`.

Trees The most dominant case is the single-column, threaded message board with indented nested replies. Usenet, Slashdot, blogs, Q/A sites, Idea platforms, all of these are basically elaborations upon this vertically-oriented layout. Some research has tried to improve upon the visualization of long threaded conversations, such as Wattenberg's focus+context interfaces [98].

Graphs Another method is to visualize utterances as a graph. This is the technique that argument visualization researchers have been fixated upon. See Figure 4.3 for a web-based example. In my opinion, this is an overexplored dead end. Ironically, considering that these approaches seek to support collective sensemaking, I have yet to use a graph-based layout that is easy to navigate or grasp any sense of the larger pattern of thoughts people are collectively expressing. Perhaps there are limited applications for personal mind-mapping activities, but their extension to multiple users has thus far appeared to be fraught with difficulties. Researchers should stop being so fixated on this technique and experiment with other approaches.

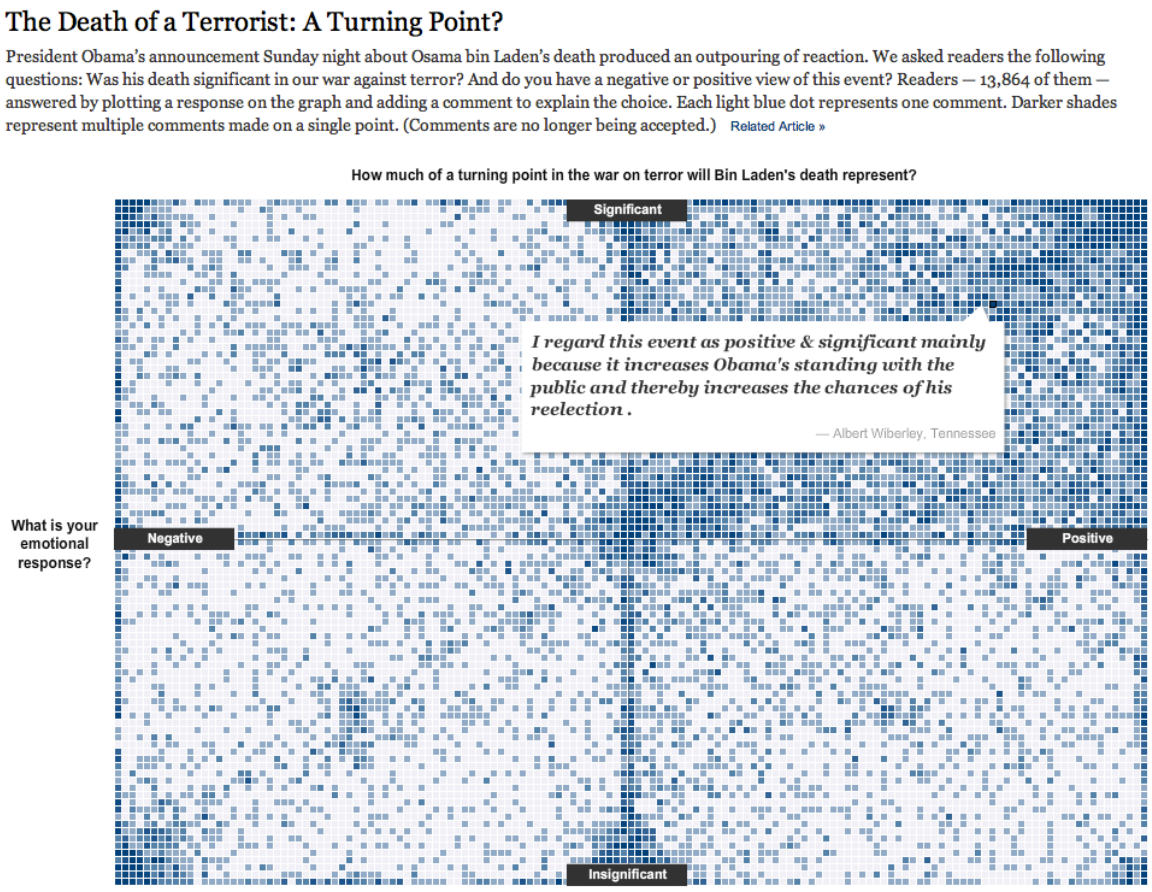


Figure 4.4: 2d visual presentation of comments, from `http://www.nytimes.com/interactive/2011/05/03/us/20110503-osama-response.html`.

2d plots OpinionSpace offers a promising new way of visualizing comments by plotting comments in a two-dimensional space (an abstract dimensional reduction based on commenters' answers to a value questionnaire). A similar technique, without dimensional reduction, was recently used by the New York Times (Figure 4.4). I reviewed more details of OpinionSpace earlier in Section 1.1.3.

Horizontal layout In both ConsiderIt and Reflect, I have experimented with layouts that exploit the horizontal dimension as a way of organizing discussions and drawing attention to different semantic types of utterances. As a whole, Reflect and ConsiderIt suggest that structuring discursive interfaces to leverage the horizontal dimensions is a productive line of inquiry for exploring further.

Reflect introduces a second dimension into a comment board in order to carve out a backchannel for listening. Interpretations of meaning on the right, replies below. It is a new way of organizing web discussion to create a more balanced experience between speaking and listening, and to introduce an explicit space to seek objective understanding ("what did you mean?") in a typically subjective space ("what should we do about this issue?"). In my opinion, the horizontal layout has been an overall success for Reflect, but some challenges remain to be addressed. On one hand, the interpretations need to go on the side to distinguish interpretations from replies; if they were placed below the comments, the sensemaking affordances of side-by-side comparisons through the highlight functionality would be undermined and readers would not be able to as easily experience one iteration of grounding. However, the current design is ambiguous about whether the comment or the reflections should be read first, which causes a dissonant reading experience. I still have not figured out a satisfactory solution, but I believe that the Reflect column needs to somehow be made secondary to make it clear that the comment should be read first, then the bullet points (which is my tentative conclusion at this point).

ConsiderIt similarly exploits the horizontal dimension for encouraging reflective discussion. First, the two-column pro/con list format highlights imbalance between pros and cons, nudging users to consider tradeoffs, or at least acknowledge when they are not. Second, additional columns on either side of the pro/con list contain the pros and cons authored by others. This four-column layout structurally nudges people to consider the perspective of other people. As discussed earlier, this nudge is likely to be most strong when users are puzzling over what a con point would be for a decision they support, or vice versa. Thus, the structural nudge may be strongest when it has the most potential to support engagement across political divides. The four column layout thus structurally encourages two distinct types of reflective behavior in a usable and simple way. Moreover, it supports a UX that explicitly focuses attention on personal deliberation (the central two columns are one's own list), while being able to secondarily engage others' points (outer columns), helping to hierarchically order attention and deemphasize confrontational encounters with others. Finally, without the side-by-side column layout, the point inclusion functionality, the crux of ConsiderIt, would be difficult if not impossible to convey in the design.

To generalize from these lessons about how the horizontal layout can provide leverage for supporting reflective public discussions, I highlight three affordances:

1. *Understandable*. By creating different, fixed spaces for utterances serving different purposes, the high-level structure of a discussion space can be easily conveyed (ConsiderIt: pro/con format, others' thoughts in the margin; Reflect: first order comments on the left, listening backchannel on the right).

2. *Actionable*. Reflective actions can easily be conveyed by making transitions between the spaces apparent (ConsiderIt: point inclusions as arrow from margin into one's personal pro/con list; Reflect: hover bullet + highlight comment, adding a bullet point off to the right).

3. *Enforceable*. By delimiting spaces as being intended for certain types of utterances, it becomes easier for a community to moderate it. This is particularly true for Reflect, where there is a more justifiable response to accusations of censorship for moderating this space, as compared to moderating the first-order comments. This affordance can be framed as creating a honeypot for attracting trolls who can now be sanctioned on a more legitimate basis for violating norms.

Chapter 5

CONCLUSION AND FUTURE WORK

By building interfaces that subtly encourage people to consider issues, listen to the diverse perspectives of others, and introspect about their own beliefs, I argue that we can help build public trust while improving upon our collective ability to address systemic problems in our institutions and ultimately take more effective action on problems such as financial reform and climate change.

In the remainder of this section, I will first discuss future directions for innovation in supporting public discussion on the web, starting from Reflect and ConsiderIt individually, but ultimately discussing higher level participatory governance workflows that use these systems as building blocks toward collective intelligence. I will then conclude by summarizing my intellectual contributions toward this goal.

5.1 Future work: toward robust collective intelligence

While organizations are exploring new ways of involving interested constituents in decision-making, these efforts are often overwhelmed by the sheer scale of the discussion. Governments in particular often lack the resources necessary to extract value from the discussion and respond. This is the problematic division of communicative labor I discussed earlier: people speak individually, proliferating their comments, and then at the end, representatives from the organization may in some cases attempt a resource-intensive summarization process.

I believe we need to reformulate this division of labor in order to create the conditions under which we can intelligently confront the problems we all face. Specifically, we should empower communicators as a group to collaboratively distill, link together, relate and summarize their thoughts (and make them responsible for doing so as part of the bargain of being part of a decision making process). If we can create the social infrastructure for participants to summarize their thoughts, we can enable decision makers and other participants to actually engage with the points being expressed in a large discussion. A wide range of high-value applications would be opened. For example, a news site might incorporate points raised in a discussion into their decision about followup stories. Or a city government might host a series of discussions about a proposed policy, incorporating feedback at various key points in its lifecycle and showing how the feedback impacted the bill. Or a grass roots movement might host discussions to establish a movement's priorities and foster a shared identify and purpose.

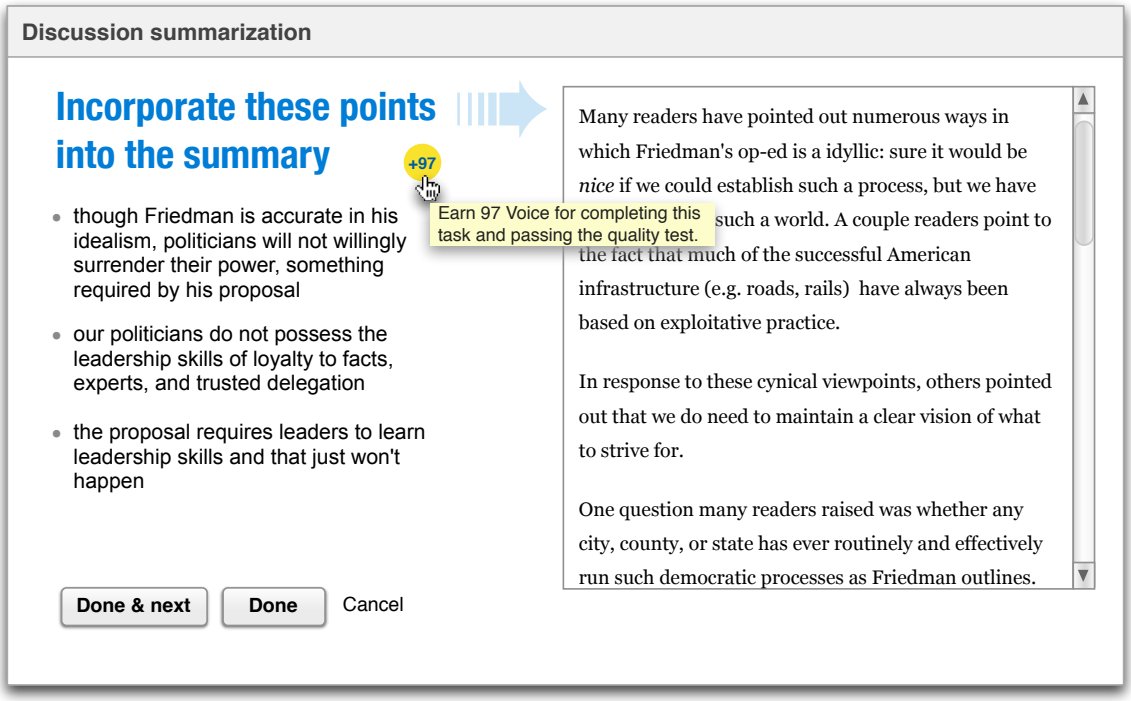


Figure 5.1: One possible task in a summarization workflow. Summarizer is asked to incorporate Reflect restatements into a partially created discussion summary.

While Reflect and ConsiderIt do not directly tackle the reformulation of this division of labor, they do provide a strong foundation for integrating, linking and summarizing the points many people are making. Both Reflect and ConsiderIt provide leverage for summarization (e.g. comment-level summaries in Reflect and PointRank in ConsiderIt), but they can be extended to support more integrative summaries – summaries that distill meaning across statements made by multiple people. In the rest of this section, I describe extensions of Reflect and ConsiderIt that move toward integrative summarization, and then discuss cross-platform / cross-context governance applications that use Reflect and ConsiderIt as building blocks.

5.1.1 Reflect

Ironically, Reflect actually produces more text than just the comments, far from creating succinct discussion-level summaries. But, as the Wikimedia deployment showed, Reflect restatements can

act as comment-level summaries useful in writing discussion-level summaries. I believe that it would be fruitful to use the restatements to fuel a participant-powered discussion summarization activity that yields a Wikipedia-like summary of the discussion. The summary would capture the most important takeaways from the discussion. It could be gradually written over time by many readers doing small tasks that are automatically combined, such as incorporating a set of restatements into the summary, verifying that a restated point is reflected in the summary, and improving the flow of the summary (see e.g. Figure 5.1). One major research question is whether bulleted Reflect points can be transplanted into a summarization context; it may be the case that they are too decontextualized to be of use during that process.

While some participants might enjoy the summarization activity, the motivation to participate in the discussion-level summarization activity may fall short. Reflect might again provide leverage. Future work could investigate a discussion-oriented reputation system where users need to "spend" characters in order to leave their comments. Users would earn the right to speak more not just by speaking well (judged via social voting), but also by listening well through restatements that other users can evaluate, creating grist for summarization. Moreover, users could earn more currency by completing discussion-level summarization tasks. Such a system would thus enable a social organization that promotes a balance between speaking and listening, as well as an obligation to help create synthesize and summarize.

This would be the full Reflect discussion system: the summarization functionality, combined with a Reflect enabled discussion board and reputation system that rewards speaking well, listening attentively and working to synthesize the current discussion.

5.1.2 ConsiderIt

Although participants appeared to reduce their rate of contributing pros and cons over the course of the five week Living Voters Guide deployment, participants' pros and cons often overlapped in sentiment and substance, creating redundancy that distracted participants and harmed the LVG's ultimate utility as an informational guide that could shine light on different facets of a ballot measure. If we are to scale ConsiderIt deployments nationally or in other high-volume contexts, we need to innovate with new approaches for managing redundancy. Eliminating redundancy is not an option,

as self-expression is an important aspect of political discourse.

In ConsiderIt, I believe the way forward is to cluster points thematically via crowdsourcing and/or NLP, building a network of points linked by the semantic distance between them (a similarity graph). With a tractable approach to labeling the semantic distance between points, a number of opportunities for managing redundancy are opened.

First, a similarity graph would allow us to better select points to show during position crafting, such that the set as a whole speaks to a diversity of facets of the issue, rather a potentially redundant set of points.

Second, a similarity graph would help identify and facilitate opportunities for summarization. One possibility is to contact the authors of two candidate redundant points and ask them to come up with a joint point that merges the two redundant ones. Not only might this lead to more thoughtful points, but the participants might gain a sense of increased efficacy in their ability to communicate and work with fellow citizens. A second possibility is to foster a collaborative process of summarizing points related to particular themes identified through the clustering. For example, a budgetary ballot measure might impact the education system, and participants might create pros and cons that relate to that educational impact. A summary statement that succinctly weaves together the various pros and cons relating to education could make for a valuable guide to that aspect of the decision.

Clustering points is not the only starting point for moving toward more succinct, integrated statements made by multiple people. Another possibility is to leverage the fact that ConsiderIt captures participants' stance on the issue and then enables them to explore the pros and cons that other voter groups found to be most important. What if after submitting a pro/con list and a stance, users had the opportunity to contribute to a collaboratively written position statement crafted by those who took that stance? Other participants who are exploring why a group took that particular stance could then read an integrated account.

5.1.3 Robust collective intelligence

There are trajectories to decision-making: identifying and defining a problem, brainstorming about what can be done, identifying concrete approaches, narrowing options and deliberating about their merits, and ultimately choosing an option and implementing it. The decision is often a product of

discussions with many stakeholders over a long period criss-crossing these different orientations. And for networked publics, the impetus to deliberate about taking some action on a public issue often only arises after many open-ended discussions that start to highlight that there is any public issue at all.

ConsiderIt and Reflect, along with other specialized discussion technology, are useful during different phases of participatory decision making.

First, there can be ephemeral discussions which are only tied to decision making in the loose sense that they can impact the formation of personal/public opinion. Discussions rooted to news stories, discussions raging across the blogosphere, and conversations on Facebook and Twitter all fall into this category. Open ended discussion of current events, in the absence of any immediate decision to be made, is valuable for sensitizing publics to collective problems, potential ways of overcoming them, and fostering understanding of what other people think. The Slashdot deployment of Reflect is an example of this type of conversation. Summarizing these kinds of discussions would be useful for starting to identify emergent issues that publics are grappling with, and their initial understanding of those issues.

Second, there can be brainstorming about what to do about an issue facing a community. Idea platforms are particularly useful during this type of deliberation, as their limited formalism can help focus the discussion slightly. However, in some cases, open-ended discussion boards are perhaps still preferred, as they do not constrain users in what they can say as they grapple with the general issue. A Reflect-enabled comment board would be useful in these conditions as a mechanism for identifying unique solutions, or interesting problem framings. The Mechanical Turk deployment of Reflect is a case in point, although I could have phrased the Reflect prompt differently to something like “restate a specific idea [user] has for Amazon”.

Third, there can be non-binding deliberations about a specific decision to be made. ConsiderIt is designed for this case, where a question can be posed that has a response that can be answered along a spectrum. I call this a closed-prompt deliberation (“Should we do this?”), as opposed to open-prompts (“what can we do?”). Because the decision is not binding, it is primarily useful for helping other members of the public understand what their peers believe, and, if relevant, informing decision makers about what what its constituents think. It can be a tool for identifying emerging consensus, although it does not need to be in order to be useful.

Fourth, there can be closed-prompt discussions where the votes that are cast are tallied and treated as binding for the given decision. Because of security problems (e.g. sockpuppets), I do not believe that we are at the point where web-based binding discussions are feasible, at least in the political realm. Regardless, systems like ConsiderIt can also support this type of discussion.

There has been little, if any, research on the efficacy of different governance workflows across these different phases, or any research seeking to create more seamless transitions between the underlying technologies. Instead we have fairly unsuccessful efforts like Obama's Dialogue on Open Government, which started with an idea system, transitioned in a second phase to multiple blog postings, then finally into a confusing wiki-based system for authoring policy. At each step, participation dropped off precipitously, and I believe at least part of that was because each stage was technologically isolated from the prior step – no one could see how their participation at the prior step impacted the next step.

Experimenting with different workflows and interactive systems, and different ways of transitioning between these phases, will be a high-impact research program. One good place to start is with the participatory governance experiment I started on Mechanical Turk. Mechanical Turk would be ideal for this kind of governance experimentation because Mechanical Turk itself is equipped to naturally handle A/B experiments. The A/B Reflect/no-Reflect experiment basically kickstarted what could be a large, multi-phase experiment in catalyzing the Mechanical Turk community to organize around their working conditions. We could, for example, take the long conversations we started and run a series of experiments for summarizing the discussion, focusing on actionable ideas. The resulting ideas identified through summarization could then be fed into an IdeaScale instance, where Turkers would be invited to vote and comment on the ideas they believed had the most traction. The most promising of these ideas could then transition into a ConsiderIt instance, where Turkers would be invited to deliberate about the relative merits of each. The ideas getting the most support would then become the proposed candidate changes for Amazon to act on. We hold an election on Mechanical Turk for electing a Seattle-based Turker to present each idea to the Mechanical Turk team at Amazon (a meeting which we would facilitate).

Ultimately, this line of research moves more toward the vision laid out in the introduction: with an attention to the full decision-making substrate, we will have more leverage in experimenting with different institutional designs that better embed the conversation in a meaningful context. This will

allow us to better understand how individual and collective efficacy is impacted by the sociotechnical design, as well as evaluate the quality of the decisions being made under different regimes.

5.2 Intellectual contribution

The most important contribution is the conception, design and engineering of ConsiderIt and Reflect, two novel systems that centralize reflective mechanisms, thus supporting activities like listening that are commonly ignored in academic and industrial systems. Reflect and ConsiderIt help demonstrate that it is possible to design lightweight platforms for public discussion that hinge upon participants reflecting upon points that other participants have raised. The *motivating thesis* for these reflective systems is that:

> Web interfaces that (1) nudge people to reflect on the issue and/or other participants' thoughts and (2) re-present evidence of this reflection can improve participants' (a) ability to find satisfaction in the discussion and (b) ability to make sense of what other people are saying, enabling publics to build upon each others' thoughts to solve common problems.

My empirical work is not controlled enough to claim that I have *demonstrated* the above motivating thesis; whether the act of using these reflective mechanisms led to the theorized outcomes, and whether those outcomes would improve the ability for groups to solve problems, remains an open question. For Reflect, there is some anecdotal evidence from the discussion facilitation interviews that Reflect helps restaters better make sense of what others are saying and potentially improve satisfaction; on the other hand, self-report data from our controlled Mechanical Turk study showed no clear results in distinguishing a Reflect-enabled discussion from a non-Reflect enabled discussion in terms of how participants perceived each other and the productivity of the discussion. For the ConsiderIt deployments, I have no controlled experiments.

Instead, I have demonstrated empirically that *the general public will voluntarily participate, without prompting, in these systems* through a series of strong field deployments of Reflect and ConsiderIt. I have deployed Reflect in three different contexts, in particular with Slashdot, where we saw a great deal of use of the restatement functionality. I have deployed ConsiderIt for use

in two state elections, each time with significant traffic and use of the reflective mechanisms. In the first ConsiderIt deployment, users spent more than ten minutes on average at the LVG, many of them returning multiple times. I believe that this should be first step of any project seeking to impact public discussion and deliberation. Too many civic systems designed in the academy seek to build a utopia without engaging human motivation. 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, I have shown that users generally appropriated these systems in manners consistent with the designed intent, using the reflective mechanisms in normatively desirable ways. For Reflect, all bullet points written in the discussion facilitation contexts were fairly neutral restatements of the respective commenter. But even in Slashdot, a community not known for its empathetic listening, over half of the bullets created reflected meaning, with community moderation reinforcing the use of Reflect as listening functionality, where the community sanctioned bullets that did not reflect meaning. With the first Living Voters Guide deployment, we found that ConsiderIt users frequently crafted pro/con lists that acknowledged tradeoffs (41.4% of all positions) and recognized points written by someone that they disagreed with (33.7% of all positions). This is a lower bound on considering tradeoffs and engaging difference: our data about whether users sought to at least read both pros and cons as well as points written by the other side shows a significantly higher frequency. This leads to what I claim as my humbler *demonstrated thesis*:

> Web interfaces that centralize the use of mechanisms encouraging participants to reflect on the issue and/or other participants' thoughts are highly engaging for general use and participants will frequently use them in normatively desirable ways.

Appendix A

INTERVIEW DATA

A.1 Based on your experience writing Reflect bullet points...

A.1.1 When was creating a Reflect bullet point useful? What did you see in a comment or commenter that made you decide to write a bullet point?

John As administrator I wanted to model the use of Reflected bullets – so I tried to always add a reflection as comments appear.

Philippe It tended to be when an author made multiple points in one post that weren't easily apparent from skimming, or when a point was particularly good and I didn't want to lose it.

Eugene Again, my orientation was around summarizing threads, so I would try to pull out key points in Reflect bullet points.

Kristoffer Some discussions in the strategic planning process became very lengthy, and therefore there was a large amount of text anyone that not had been active from the start of a particular thread had to read before he/she was able to participate constructively in there. In addition the discussions had to be condensed into much shorter conclusions that could fit in a final report. So the main reason I wrote bullet points was to address these two issues; to make new participants able to quickly join discussions, and to help summarize the discussion into shorter statements.

A.1.2 What makes a good Reflect bullet point? What goes through your head when you are creating a Reflect bullet point?

John Short and sweet! Trying to write the least I can to capture the essence of a small portion of the original document.

Philippe It's almost like writing a tweet: short, to the point, and provided the necessary info to recreate my thought process later. (Any historical data, etc).

Eugene My goal was always to highlight the essence of the post, no more, no less. I'd try to do it in a way that captured the tone of the original post. I didn't want to project my own interpretations onto the summary.

Kristoffer I tried to identify single coherent thoughts within one or several paragraph and summarize the thought in a single sentence. In the discussions statements often were backed up with context, examples and facts which often made for lengthy comments. What was good with reflect was that it was possible to strip away all the "proofs" of the statement and express the idea in a single

sentence, but still having the text that the bullet was derived from highlighted when a user hovered it and thus making it possible for anyone critical of the statement to investigate the reasoning more thoroughly.

A.1.3 When was it easy to restate a point that someone was making? When was it difficult? Do you think it's more valuable when it's difficult? Why?

John It is easier if there is more content in the original comment. It is harder when there is less material originally posted. I thing it would be impossible to Reflect if the original poster makes a short, flippant, comment. In contrast, when there is more original comment to Reflect on, it is easier – and more valuable (I think) to both the original poster and other readers.

Philippe For me, it's much more valuable when it's difficult: the easy comments didn't get reflect points, in my workflow. I only used them when I was attempting to boil down someone's thoughts. I found myself using them on non-native English speakers more than other times.

Eugene I didn't feel a lot of anxiety about restating people's points, thanks to the confirmation feature.

Kristoffer It was most difficult when it was difficult to understand what the commenter wanted to say with his/her comment. And I really think that in these cases the bullets served a purpose in forcing me to really try to understand. When you force yourself to summarize what you read, it's not possible to just skip a paragraph you don't understand without noticing that you actually didn't understand it. Something I think happens very often otherwise. But which is more important, I don't know. The passages that was easier to summarize maybe were better thought through to begin with and was more valuable in that sense. At the same time as the bullet points were valuable in the sense that they forced you to really try to understand parts you otherwise would have skipped.

A.1.4 In trying to summarize a comment, did you ever come to a better understanding of what they were trying to say?

John Not so much a "better understanding" but another way to explain the thought in a more direct way. Condensing the original content – re-packing it instead of un-packing it.

Philippe Frequently.

Eugene It definitely forced me to read more carefully. Occasionally, I would read a sentence that I didn't completely understand, and I would generally just gloss over those. You can't do that when you're looking to summarize.

Kristoffer Yes, indeed. I also think it made me more aware of the value of writing clearly in the first place.

A.1.5 Who was the primary audience for the bullet points you wrote, and what benefits would you expect them to receive? One Reflect user said that the primary audience is the author of the comment, with the benefit of demonstrating that she heard the comment author. Another Reflect user said that the audience is other readers, who would have the benefit of an easily-perused summary. Yet another said that the primary audience is himself – that writing the point helps him digest and understand the material. Which view is closer to your own – or if none of them, what is your view?

John When I Reflect, I'm doing it with a sense that I want to engage the unseen readers. It's not so very much for the original author – except to try and let him/her know that someone is reading and responding to his/her comments.

Philippe For me the primary audience is myself, for review purposes. A secondary audience is other readers. I was less concerned with the author of the post, though it did several times clarify misunderstandings.

Eugene For me, definitely other readers. This might have been different had Reflect been deployed as part of the system and not as an optional extension.

Kristoffer As mentioned in the first question, the audience I thought of was latecomers to the discussion that would be scared to enter the discussion because of the massive amount of text they would have to read through first, as well as participators in the discussions that later wanted to draw conclusions from the discussion without having to reread everything.

A.2 Based on your experience having a comment you wrote Reflected by someone else...

JS and K skipped this question group because they never had their comment summarized

A.2.1 What was your reaction to having someone restate what they thought they heard you say?

Philippe I checked to be certain they had it correct. I also wondered if I possibly could have phrased something more clearly. I found myself editing comments afterwards for style (removing subordinate clauses, difficult punctuation, etc).

Eugene I loved it! First of all, it meant that someone else was using Reflect. But it was also explicit validation that someone understood what I was talking about. On occasion, it was humbling, as people sometimes expressed my point much more concisely.

Did it feel different than if someone simply addressed your point in a subsequent comment?

Philippe Yes, but I think that's a function of the fact that it was a new tool.

Eugene Yes.

A.2.2 Reflect lets you confirm whether a restatement of your comment was accurate or not, and leave a clarifying remark if not. Did you use this functionality? What do you think about this functionality?

Philippe Yes, and I like it a great deal. That was a good way to tell whether the original author approved.

Eugene Yes. Loved it. I think it's a key feature, because the one thing you fear – both as an author and as a summarizer – is being misrepresented or misrepresenting others. This provided a safety net.

A.3 Based on your experience reading a Reflect-enabled discussion...

A.3.1 Was it useful to hover over a Reflect bullet point and see the relevant text highlighted in the comment?

John Yes, very muchâ it helped to focus my attention and I don't have to Reflect on the entire comment but only on a specific portion.

Philippe Absolutely. Sometimes it was difficult to highlight the relevant text (picking it out, etc) but when reading, this was important functionality.

Eugene Yes.

Kristoffer This was one of the features I really liked about Reflect. Sometimes I wrote a bullet point that I couldn't highlight, and that often meant that the bullet point wasn't all too accurately formulated, and I had to reconsider if I actually had understood what the reader really meant yet. As mentioned in the first question set it also served an important role in making it clear how a certain statement was backed up with additional examples, context, and facts in the actual comment.

A.3.2 Did having comments with restatements on the right change how you read comments? If so, how?

John Not enough Restatements from others for me to answer.

Philippe I found myself scanning the comments first to get an idea what to look for when reading the comments. This was both good and bad.

Eugene Yes. It made it much easier for me to skim and bear down on posts that interested me.

Kristoffer I did mostly write bullet points, and did not read enough of them to be able to give any answer to this question.

A.3.3 Did you ever come to better understand what a commenter was saying by reading someone else's summary?

Philippe Yes.

Eugene Hard to say. If I had to guess, I'd have to say no, although it may have more to do with the nature of the conversation and the limited usage of Reflect on our site.

Kristoffer I didn't read very many comments that had bullet points and can unfortunately not tell if that is the case or not.

A.3.4 Did you ever jump into the middle of a Reflect-enabled discussion, or need to catch up on one after being away from it for a little while? If so, how did the Reflect summaries affect your experience?

Philippe No, because I was one of the main people driving it.

Eugene Absolutely. Reflect summaries definitely made it easier to catch up on threads.

Kristoffer No.

A.4 Based on your experience participating in and possibly leading community discussions (not necessarily with Reflect)...

A.4.1 What other kinds of discussions do you think Reflect would be good for? Why? What kinds of discussions would Reflect NOT be good for? Why?

John Honestly, I think Reflect would be good for all types of discussions. Just having the invitation to "slow down and reflect" on what is being said is an important service.

Philippe [It would be good for] Balloting/voting, for sure - particularly in a Request for Comment format. Anything that required a complex thought would tend to get distilled into a "comic" summation of what it actually is, I think.

Eugene On a theoretical level, I think that Reflect would be great for any discussion thread with long posts. On a practical level, you'd need a community of readers who might be compelled to actually do the summaries. It would be useless on my blog, given that so few people post there; any discussion where posts are generally short, perhaps more informational.

Kristoffer I think any discussions that can become lengthy can benefit from Reflect. I think either the case where a few active participants wants to be able to summarize their discussion and draw conclusions from it, or the case where many participants are supposed to be able to follow what a few have discussed can benefit from it. I also often find myself reading forum posts on internet when I google for something. maybe many forum boards could benefit their visitors by having similar capabilities. I have recently participated in conversations on TED.com, and there discussions tend to be very lengthy. Maybe that's a place that would be willing to implement such functions. Also, In discussions where it is difficult to formulate yourself well because the concepts are difficult, I think Reflect can offer an insight into if you succeeded in getting your point across. It would not be good on discussions where answers anyway are short. There is no need for Reflect on Facebook, at least on the pages I have visited.

A.4.2 What impact on the quality of discussion do you think Reflect might have?

John Asking the reader to pause – even if he/she does not actively use Reflect – and even "think" about Reflecting would, I think, be important for a discussion thread.

Philippe I think that it's particularly important in conversations where some participants are

not native English speakers. It allows for a confirmation of the point they're trying to make in a non-threatening and non-judgmental way.

Eugene I think it definitely improves the reader experience and the ability for people to comprehend complex discussions. I also think that any tool that encourages listening and provides that feedback will have a positive impact both on the quality of the content as well as on the relationships of the people having the discussion. That can only be beneficial.

Kristoffer It makes it easier for people to identify which large chunks of text that are worth investing time in. There is so much text on the internet, and being able to identify which information contains valuable information without reading the text before would helpful.

A.4.3 What impact on participants' attitudes toward other participants do you think Reflect might have?

John Reflect asks people to "stand a moment in the shoes of another" – to pause and really listen to what the other person is trying to say. This is how I think Reflect improves a discussion.

Philippe I think it forces a slowdown and pause. We tend to jump instinctively on comment threads and Reflect forced me to stop and consider the underlying meaning of what the commenter said, as opposed to just what I initially read.

Eugene From the point of view of an author, I think knowing that other people are listening to you definitely has a positive impact. I'm not sure that it will have much of an impact on the person doing the summarizing.

Kristoffer Don't really know.

A.4.4 If you could change one thing about Reflect, what would it be and why?

John Allow Reflect to use audio reflections (moderated of course) and limit the audio reflection to something short – maybe just 20 seconds? Also, visually show how often the person writing the Reflection has written other Reflections? Perhaps adding a summary list to the person's collection of other Reflections?

Philippe The implementation that we used was not integrated into the interface. That made things difficult. In a perfect world, there'd be no need for greasemonkey scripts, etc.

Eugene I can't think of anything right now. You were very responsive to suggestions when we were using the tool. One of the things that really appealed to me about Reflect was that the design was light, but the thinking underlying the design was deep.

Kristoffer It was slow to load when I used it. I had to explicitly allow the script to run to completion each time I loaded a page. I would definitely like that to change.

BIBLIOGRAPHY

[1] Mark S. Ackerman, Anne Swenson, Stephen Cotterill, and Kurtis DeMaagd. *I-DIAG: from community discussion to knowledge distillation*, pages 307–325. Kluwer, B.V., 2003.

[2] B. Thomas Adler, Krishnendu Chatterjee, Luca de Alfaro, Marco Faella, Ian Pye, and Vishwanath Raman. Assigning trust to Wikipedia content. In *Proceedings of the 4th International Symposium on Wikis*, WikiSym '08, pages 26:1–26:12, New York, NY, USA, 2008. ACM.

[3] Mei Alonzo and Milam Aiken. Flaming in electronic communication. *Decision Support Systems*, 36(3), 2004.

[4] Christopher Anderson. Deliberative, agonistic, and algorithmic audiences: Journalism's vision of its public in an age of audience transparency. *Int. J. of Communication*, 5, 2011.

[5] Jaime Arguello, Brian S. Butler, Elisabeth Joyce, Robert Kraut, Kimberly S. Ling, Carolyn Rosé, and Xiaoqing Wang. Talk to me: foundations for successful individual-group interactions in online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 959–968, New York, NY, USA, 2006. ACM.

[6] Brian P. Bailey and Eric Horvitz. What's your idea?: a case study of a grassroots innovation pipeline within a large software company. In *Proceedings of the 28th International Conference on Human factors in Computing Systems*, CHI '10, pages 2065–2074, New York, NY, USA, 2010. ACM.

[7] Eric P.S. Baumer, Jordan Sinclair, and Bill Tomlinson. America is like metamucil: fostering critical and creative thinking about metaphor in political blogs. In *Proceedings of the 28th International Conference on Human factors in Computing Systems*, CHI '10, pages 1437–1446, New York, NY, USA, 2010. ACM.

[8] Yochai Benkler. *The Wealth of Networks*. Yale University Press, 2006.

[9] W. Lance Bennett. Changing citizenship in the digital age. *The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning*, pages 1–24, December 2007.

[10] Ivan Beschastnikh, Travis Kriplean, and David McDonald. Wikipedian Self-Governance in Action: Motivating the Policy Lens. In *International Conference on Weblogs and Social Media*, 2008.

[11] Matt Billings and Leon A. Watts. Understanding dispute resolution online: using text to reflect personal and substantive issues in conflict. In *Proceedings of the 28th International Conference on Human factors in Computing Systems*, CHI '10, pages 1447–1456, New York, NY, USA, 2010. ACM.

[12] Alan Borning, Batya Friedman, Janet L. Davis, Brian T. Gill, Peter H. Kahn, Jr., Travis Kriplean, and Peyina Lin. Public participation and advocacy in information design. *Info. Polity*, 14(1,2):61–74, 2009.

[13] A. J. Bernheim Brush, David Bargeron, Jonathan Grudin, Alan Borning, and Anoop Gupta. Supporting interaction outside of class: anchored discussions vs. discussion boards. In *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community*, CSCL '02, pages 425–434. International Society of the Learning Sciences, 2002.

[14] Simon Buckingham Shum. Cohere: Towards web 2.0 argumentation. In *Proceedings of the 2008 Conference on Computational Models of Argument: Proceedings of COMMA 2008*, pages 97–108, Amsterdam, The Netherlands, The Netherlands, 2008. IOS Press.

[15] Heather Campbell and Robert Marshall. Public involvement and planning: Looking beyond the one to the many. *International Planning Studies*, 5(3), 2000.

[16] Joseph Cappella, Vincent Price, and Lilach Nir. Argument repertoire as a reliable and valid measure of opinion quality: Electronic dialogue during campaign 2000. *Political Communication*, 19(1):73–93, 2002.

[17] Chad Carr. *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, chapter Using Computer Supported Argument Visualization to Teach Legal Argumentation. Springer, 2003.

[18] Elizabeth F. Churchill, Jonathan Trevor, Sara Bly, Les Nelson, and Davor Cubranic. Anchored conversations: chatting in the context of a document. In *Proceedings of the SIGCHI Sonference on Human factors in Somputing Systems*, CHI '00, pages 454–461, New York, NY, USA, 2000. ACM.

[19] Hebert H. Clark and S. A. Brennan. Grounding in communication. In L. B. Resnick, J. M. Levine, and S. D. Teasley, editors, *Perspectives on socially shared cognition*. American Psychological Association, 1991.

[20] Jeff Conklin. *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, chapter Dialog Mapping: Reflections on an Industrial Strength Case Study. Springer, 2003.

[21] Jeff Conklin and Michael L. Begeman. gIBIS: a hypertext tool for exploratory policy discussion. In *Proceedings of the 1988 ACM Conference on Computer-supported Cooperative Work*, CSCW '88, pages 140–152, New York, NY, USA, 1988. ACM.

[22] Jeff Conklin, Albert Selvin, Simon Shum, and Maarten Sierhuis. Facilitated hypertext for collective sensemaking: 15 years on from gIBIS. *HYPERTEXT '01: Proceedings of the 12th ACM Conference on Hypertext and Hypermedia*, 2001.

[23] Jeff Conklin, Albert Selvin, Simon Buckingham Shum, and Maarten Sierhuis. Facilitated hypertext for collective sensemaking: 15 years on from gIBIS. In *Proceedings of the 12th ACM Conference on Hypertext and Hypermedia*, HYPERTEXT '01, pages 123–124, New York, NY, USA, 2001. ACM.

[24] Gregorio Convertino, Helena M. Mentis, Mary Beth Rosson, Aleksandra Slavkovic, and John M. Carroll. Supporting content and process common ground in computer-supported teamwork. In *Proceedings of the 27th International Conference on Human factors in Computing Systems*, CHI '09, pages 2339–2348, New York, NY, USA, 2009. ACM.

[25] Lincoln Dahlberg. Democracy via cyberspace. *New Media & Society*, 3(2), 2001.

[26] Nicholas Diakopoulos, Sergio Goldenberg, and Irfan Essa. Videolyzer: quality analysis of online informational video for bloggers and journalists. In *Proceedings of the 27th International Conference on Human factors in Computing Systems*, CHI '09, pages 799–808, New York, NY, USA, 2009. ACM.

[27] Nicholas Diakopoulos and Mor Naaman. Towards quality discourse in online news comments. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, CSCW '11, pages 133–142, New York, NY, USA, 2011. ACM.

[28] Margaret Dickey-Kurdziolek, Matthew Schaefer, Deborah Tatar, and Ian P. Renga. Lessons from thoughtswap-ing: increasing participants' coordinative agency in facilitated discussions. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW '10, pages 81–90, New York, NY, USA, 2010. ACM.

[29] John Dryzek. *Liberals, Critics, Contestations*. Oxford University Press, 2002.

[30] Starkey Duncan. On the structure of speaker-auditor interaction during speaking turns. *Language in Society*, 3(2):161–180, 1974.

[31] Thomas Erickson and Wendy A. Kellogg. Social translucence: an approach to designing systems that support social processes. *ACM Trans. Comput.-Hum. Interact.*, 7:59–83, March 2000.

[32] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. Opinionspace: a scalable tool for browsing online comments. In *Proceedings of the 28th International Conference on Human factors in Computing Systems*, CHI '10, pages 1175–1184, New York, NY, USA, 2010. ACM.

[33] James Fishkin. Strategies of Public Consultation. *Integrated Assessment*, 6(2), 2006.

[34] Andrea Forte and Amy Bruckman. Scaling consensus: Increasing decentralization in wikipedia governance. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, HICSS '08, pages 157–, Washington, DC, USA, 2008. IEEE Computer Society.

[35] Benjamin Franklin. *Mr. Franklin: A Selection from His Personal Letters*. Yale University Press, 1954.

[36] Nancy Fraser. Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text*, 25/26, 1990.

[37] Deen Freelon. Analyzing online political discussion using three models of democratic communication. *New Media & Society*, 12(7), 2010.

[38] Susan R. Fussell, Robert E. Kraut, and Jane Siegel. Coordination of communication: effects of shared visual context on collaborative work. In *CSCW '00: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, pages 21–30, New York, NY, USA, 2000. ACM.

[39] Michael Gamon, Sumit Basu, Dmitriy, Danyel Fisher, Matthew Hurst, and Arnd Christian Konig. BLEWS: Using Blogs to Provide Context for News Articles. In *International Conference on Weblogs and Social Media*, 2008.

[40] R. Kelly Garrett. Echo chambers online?: Politically motivated selective exposure among Internet news users. *J. of Computer Mediated Communication*, 14(2), 2009.

[41] John Gastil. *Political Communication and Deliberation*. Sage, 2008.

[42] Jonathan Grudin. Groupware and Social Dynamics: Eight Challenges for Developers. *Communications of the ACM*, 37(1), 1994.

[43] Jürgen Habermas. The Public Sphere: An Encyclopedia Article (1964). *New German Critique*, (3):49–55, 1974.

[44] Jürgen Habermas. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. MIT Press, 1991.

[45] Sean Hansen, Nicholas Berente, and Kalle Lyytinen. Wikipedia, critical social theory, and the possibility of rational discourse. *The Information Society*, 25:38–59, January 2009.

[46] William C. Hill, James D. Hollan, Dave Wroblewski, and Tim McCandless. Edit wear and read wear. In *CHI*, 1992.

[47] Bernardo Huberman. *The Laws of the Web*. MIT Press, 2001.

[48] J. Innes and D. Booher. Reframing public participation: Strategies for the 21st century. *Planning theory and practice*, 5(4), 2004.

[49] Severin Isenmann and Wolf D. Reuter. IBIS: a convincing concept...but a lousy instrument? In *Proceedings of the 2nd Conference on Designing interactive systems: processes, practices, methods, and techniques*, DIS '97, pages 163–172, New York, NY, USA, 1997. ACM.

[50] Shanto Iyengar and Robert Luskin. Deliberative public opinion in presidential primaries: Evidence from the online deliberative poll. In *Voice and Citizenship Conference*, 2004.

[51] Nicholas Jankowski and Renee van Os. *The Prospects for Political Renewal Through the Internet*, chapter Internet-based political discourse. Routledge, 2004.

[52] Cheryl King, Kathryn Feltey, and Bridget Susel. The question of participation: Toward authentic public participation in public administration. *Public Administration Review*, 58(4), 1998.

[53] Paul A. Kirschner, Simon J. Buckingham Shum, and Chad S. Carr, editors. *Visualizing argumentation: software tools for collaborative and educational sense-making*. Springer-Verlag, London, UK, 2003.

[54] Aniket Kittur, Ed H. Chi, and Bongwon Suh. Crowdsourcing user studies with mechanical turk. In *Proceeding of the twenty-sixth annual SIGCHI Conference on Human factors in Computing Systems*, CHI '08, pages 453–456, New York, NY, USA, 2008. ACM.

[55] Mark Klein and Luca Iandoli. Supporting collaborative deliberation using a large-scale argumentation system. In *Online Deliberation*, 2008.

[56] Silvia Knobloch-Westerwick and Jingbo Meng. Looking the other way: Selective exposure to attitude-consistent and counterattitudinal political information. *Comm. Research*, 36(3), 2009.

[57] Tetsuro Kobayashi and Kenichi Ikeda. Selective exposure in political web browsing. *Information Communication Society*, 12(6):929–953, 2009.

[58] Kenneth L. Kraemer and John Leslie King. Computer-based systems for cooperative work and group decision making. *ACM Comput. Surv.*, 20:115–146, July 1988.

[59] R E Kraut, S H Lewis, and L W Swezey. Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology*, 43(4):718–731, 1982.

[60] Travis Kriplean. Round pegs and square holes, or collaboration in citizen-citizen-government interaction. *Unpublished, see* `http://www.cs.washington.edu/homes/travis/papers/engage/com597a09.pdf`, 2009.

[61] Travis Kriplean, Ivan Beschastnikh, Alan Borning, David W. McDonald, and Mark Zachry. Mediating spaces between citizens and government. In *Socially Mediating Tech. Workshop at CHI*, 2009.

[62] Travis Kriplean, Ivan Beschastnikh, and David W. McDonald. Articulations of wikiwork: uncovering valued work in wikipedia through barnstars. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08, pages 47–56, New York, NY, USA, 2008. ACM.

[63] Travis Kriplean, Ivan Beschastnikh, David W. McDonald, and Scott A. Golder. Community, consensus, coercion, control: CS*W or how policy mediates mass participation. In *Proceedings of the 2007 International ACM Conference on Supporting group work*, GROUP '07, pages 167–176, New York, NY, USA, 2007. ACM.

[64] Namhee Kwon, Stuart W. Shulman, and Eduard Hovy. Multidimensional text analysis for erulemaking. In *Proceedings of the 2006 International Conference on Digital government research*, pages 157–166, New York, NY, USA, 2006. ACM.

[65] Namhee Kwon, Liang Zhou, Eduard Hovy, and Stuart Shulman. Identifying and classifying subjective claims. In *Proceedings of the International Conference on Digital Government Research (DGO)*, 2007.

[66] Claus Lamm, Daniel Batson, and Jean Decety. The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, 19(1):42–58, 2007.

[67] Cliff Lampe and Paul Resnick. Slash(dot) and burn: distributed moderation in a large online conversation space. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, CHI '04, pages 543–550, New York, NY, USA, 2004. ACM.

[68] Patricia Lange. Commenting on comments: Investigating responses to antagonism on YouTube. In *Society for Applied Anthropology*, 2007.

[69] Jonah Lehrer. *How We Decide*. Mariner Books, 2010.

[70] Lawrence Lessig. *Republic, Lost: How Money Corrupts Congress–and a Plan to Stop It*. Twelve, 2011.

[71] Kurt Luther, Scott Counts, Kristin B. Stecher, Aaron Hoff, and Paul Johns. Pathfinder: an online collaboration environment for citizen scientists. In *Proceedings of the 27th International Conference on Human factors in Computing Systems*, CHI '09, pages 239–248, New York, NY, USA, 2009. ACM.

[72] Michael MacKuen, Jennifer Wolak, Luke Keele, and George Marcus. Civic engagements: Resolute partisanship or reflective deliberation. *Am. J. of Pol. Sci.*, 54(2), 2010.

[73] Allan MacLean, Richard M. Young, Victoria M. E. Bellotti, and Thomas P. Moran. Questions, options, and criteria: elements of design space analysis. *Hum.-Comput. Interact.*, 6(3):201–250, 1991.

[74] Thomas W. Malone and Mark Klein. Harnessing collective intelligence to address global climate change. *Innovations*, 2(3), 2007.

[75] Peter Muhlberger. The virtual agora project: A research design for studying democratic deliberation. *J. of Public Deliberation*, 1(1), 2005.

[76] Sean A. Munson and Paul Resnick. Presenting diverse political opinions: how and how much. In *Proceedings of the 28th International Conference on Human factors in Computing Systems*, CHI '10, pages 1457–1466, New York, NY, USA, 2010. ACM.

[77] Diane Mutz. *Hearing the Other Side: Deliberative versus Participatory Democracy*. Cambridge, 2006.

[78] Kevin K. Nam and Mark S. Ackerman. Arkose: reusing informal information from online discussions. In *Proceedings of the 2007 International ACM Conference on Supporting group work*, GROUP '07, pages 137–146, New York, NY, USA, 2007. ACM.

[79] Souneil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. Newscube: delivering multiple aspects of news to mitigate media bias. In *Proceedings of the 27th International Conference on Human factors in Computing Systems*, CHI '09, pages 443–452, New York, NY, USA, 2009. ACM.

[80] Walter Parker. *Teaching Democracy: Unity and Diversity in Public Life*. Columbia, 2003.

[81] Francesca Polletta and John Lee. Is telling stories good for democracy? Rhetoric in public deliberation after 9/11. *American Sociological Review*, 71(5):699 –721, 2006.

[82] Vincent Price, Lilach Nir, and Joseph Cappella. Normative and informational influences in online political discussions. *Communication Theory*, 16(1):47–74, 2006.

[83] Alastair Renton and Ann Macintosh. Computer-supported argument maps as a policy memory. *The Information Society*, 23:125–133, March 2007.

[84] Warren Sack. Conversation Map: An interface for very large-scale conversations. *J. Manage. Inf. Syst.*, 17(3), 2001.

[85] Douglas Schuler. *Online Deliberation: Design, Research, & Practice*, chapter Online civic deliberation with E-liberate. CSLI Publications, 2009.

[86] Albert Selvin. *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, chapter Fostering Collective Intelligence: Helping Groups Use Visualized Argumentation. Springer, 2003.

[87] Bayle Shanks and Dana Dahlstrom. *Online Deliberation: Design, Research, & Practice*, chapter Parliament: A module for parliamentary procedure. CSLI Publications, 2009.

[88] Frank M. Shipman, III and Catherine C. Marshall. Formality considered harmful: Experiences, emergingthemes, and directions on the use of formal representations ininteractive systems. *Comput. Supported Coop. Work*, 8:333–352, October 1999.

[89] Stuart W. Shulman, Namhee Kwon, Eduard Hovy, and Emily Huisman. Tools for rules: Technology transfer and electronic rulemaking. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, page 200, Washington, DC, USA, 2008. IEEE Computer Society.

[90] Simon Buckingham Shum and Nick Hammond. Argumentation-based design rationale: what use at what cost? *Int. J. Hum.-Comput. Stud.*, 40:603–652, April 1994.

[91] Marc Smith and Andrew Fiore. Visualization components for persistent conversations. *CHI '01: Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, 2001.

[92] Stephen Coleman and Jay Blumler. *The Internet and Democratic Citizenship: Theory, Practice and Policy*. Cambridge, 2009.

[93] Abhay Sukumaran, Stephanie Vezich, Melanie McHugh, and Clifford Nass. Normative influences on thoughtful online participation. In *Proceedings of the 2011 annual Conference on Human factors in Computing Systems*, CHI '11, pages 3401–3410, New York, NY, USA, 2011. ACM.

[94] Cass Sunstein. *Republic.com 2.0*. Princeton Univ. Press, 2009.

[95] Dennis Thompson. Deliberative democratic theory and empirical political science. *Annual Review of Political Science*, 11, 2008.

[96] Michael Trice. *The Ethics of Emerging Media: Information, Social Norms, and New Media Technology*, chapter How We Comment on Web Journalism: A Case Study on Dialogue Found in News Articles. Continuum, 2011.

[97] Karin Wahl-Jorgensen. Letters to the editor as a forum for public deliberation. *Crit. Stud. in Media Comm.*, 18(3), 2001.

[98] Martin Wattenberg and David Millen. Conversation thumbnails for large-scale discussions. In *CHI '03 Extended Abstracts on Human factors in Computing Systems*, pages 742–743. ACM, 2003.

[99] Edward Weber and Anne Khademian. From agitation to collaboration: Clearing the air through negotiation. *Public Administration Review*, 57(5), 1997.

[100] S. Weber. *The Success of Open Source*. Harvard University Press, 2004.

[101] Sheila White. Backchannels across cultures: a study of Americans and Japanese. *Language in Society*, 18(1):59–76, 1989.

[102] Steve Whittaker. *Theories and methods in mediated communication*, volume 18, pages 1–53. New Jersey: Erlbaum, 2003.

[103] Steve Whittaker, Loren Terveen, Will Hill, and Lynn Cherny. The dynamics of mass interaction. In *CSCW '98: Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*, pages 257–264, New York, NY, USA, 1998. ACM.

[104] Scott Wright. Government-run fora: Moderation, censorship and the shadow of control. *Brit. J. of Politics and Int. Relations*, 8, 2006.

[105] Victor H Yngve. *On getting a word in edgewise*, volume 6, pages 567–578. Chicago Linguistic Society, 1970.

VITA

Travis Kriplean was born in Milwaukee, Wisconsin. At the University of Wisconsin-Madison he earned a Bachelor of Science degree in Computer Science and Sociology. In 2011 he earned a Doctor of Philosophy at the University of Washington in Computer Science and Engineering.