EXPOSURE TO POLITICAL DIVERSITY ONLINE

Sean A. Munson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in The University of Michigan
2012

Doctoral Committee:

Professor Paul J. Resnick, Chair
Assistant Professor Eytan Adar
Assistant Professor Mark W. Newman
Assistant Professor Joshua M. Pasek

ACKNOWLEDGEMENTS

This could easily be the longest chapter of the dissertation. I have many colleagues to thank for their role in the individual dissertation studies; I do so in the footnotes for each chapter. I am indebted to scores of others for their help with my research, graduate school, and my education more generally, but I need to take this space to thank several by name.

I came to Michigan after reading Paul Resnick's chapter on Building Sociotechnical Capital. That chapter convinced me that the School of Information is a community that cares not just about deeply understanding the relationship between people, information, and technology, but that also seeks to build systems that showcase technology's potential benefits for society. Since starting at Michigan, Paul has been steady in his support of both the importance of my research topics and of my learning, seeming to always know when to offer encouragement and when to offer critique or skepticism. Through many conversations and debates, we worked out research directions, system design choices, study designs, analyses, and how to communicate the results.

I am grateful to my committee members – Eytan Adar, Mark Newman, and Josh Pasek – who graciously worked with my suddenly-aggressive schedule for completing the dissertation while still providing thoughtful, considered suggestions on how to improve this work and where to go next. These suggestions will pay dividends for years to come.

Thank you, as well, to the entire community of faculty, staff, and students at the School of Information, but especially Michael Cohen, Mark Ackerman, Erin Krupka, Tom Finholt, Beth Yakel, Judy Olson, Mick McQuaid, Erik Hofer, Derek Hansen, Sue Schuon, Becky O'Brien, Ann Verhey-Henke, and Jocelyn Webber. I have also been lucky to have many wonderful collaborators and mentors outside of SI, including Sunny Consolvo,

Margie Morris, Caroline Richardson, Ed Chi, Julie Kientz, Jenny Preece, Daniel Avrahami, and Katie Siek.

Of course, my education did not begin with graduate school. This begins with my parents, Gary and Beth, who have encouraged and indulged my curiosity from the start, whether they were helping me send a battleship design to the Navy as a kindergartener or setting me up with Visual Basic and teaching me calculus so I could write a computer game. Throughout my life, I have been fortunate to have many fantastic teachers, teachers who forgot to tell their students what we could not do. Melissa Ryan, Eleanor Hart. Ann Murdoch, Michael Roche, and Ellen LeBlanc did their best to introduce me to research, though I thwarted these efforts and turned these projects into engineering ones. Yevgeniya Zastavker and Marc Vilain finally got me to do research, though it was not until I worked with Derek Hansen during my first year at Michigan that I finally understood how engineering and science could really go together.

Doing my undergraduate work at Olin was an amazing experience. Caitrin Lynch and Rob Martello helped me turn my disappointment about my own political blogging experiences into an academic question: a situation to study and try to improve upon, rather than to simply leave in my past. Lynn Andrea Stein and Mark Chang introduced me to the literature, and, more importantly, the community, of human computer interaction research and researchers. Without these experiences, I doubt that I would have found the path to the work that I came to Michigan to do and that is represented in this dissertation. Rod Crafts was a patient mentor and guide to the world of higher education; I credit our conversations for helping me seriously consider a career in academia.

I am grateful to others working on political information access and discussion online, who have challenged me to work harder and who have been invaluable sounding boards: Travis Kriplean, Souneil Park, Jonathan Morgan, Kelly Garrett, Eric Baumer, and Nick Diakopolous. Many other colleagues have been supportive friends, critical reviewers, and thoughtful provocateurs, as necessary, in Ann Arbor, online, and at conferences and

workshops: Jessica Hullman, Eytan Bakshy, Patrick Gage Kelley, Phil Ayoub, Daniel Zhou, Pedja Klasnja, and Erika Poole, especially. Other friends helped me escape from the world of research as necessary: thanks Matt & Emily Darragh, housemates at 406 (Greg Thorne, Aston Gonzalez, Jeremy Canfield, Shannon Curry, Allie Goldstein, Kirsten Howard, and Yariv Pierce), Morgan Mihok, Ariel Shaw, Brian Karrer, and Kim McCraw for those needed infusions of sanity.

PREFACE

"It is hardly possible to overrate the value … of placing human beings in contact with others dissimilar to themselves, and with modes of thought and action unlike those with which they are familiar… Such communication has always been, and particularly in the present age, one of the primary sources of progress."

JS Mill

"It may make your blood boil, your mind might not be changed. But the practice of listening to opposing views is essential for effective citizenship. It is essential for democracy."

President Barack Obama

The challenge of increasing individuals' exposure to and engagement with diverse political viewpoints is not merely an academic question for me. From my junior year of high school until my junior year of college, I wrote a political blog. It attracted some readership, enough to get me credentials to the 2004 Democratic National Convention and to be gently teased on *The Daily Show*. Shortly after the 2004 US Presidential election, though, I woke up reflecting that, through my blog, I had communicated almost exclusively with like-minded individuals. Yes, some of my readers, and I, had become a bit more fired up and perhaps a couple of additional people voted, but the discussion on the blog came nowhere near the diverse, thoughtful discussion that I had hoped to have, and so I stopped.

A desire to tackle this problem – to find ways to promote the deliberative discourse that had not materialized on my blog, or to find places where it might already

occur – motivated me to come to graduate school. This dissertation summarizes the efforts my colleagues and I have made on one small but important piece of this problem – increasing individuals' exposure to diverse points of view – over the past five years.

Like most dissertations, this one ends up being smaller than the original ambition. I hope, though, that it moves me, and our community, closer to our goals.

Sean Munson

12 July 2012

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

Introduction

Diversity in the information people see and attend to can help people and groups make better decisions, can help people learn and correct inaccurate beliefs, and can help people see ideas with which they do not agree as legitimate. Media policy in the United States has long had a focus on promoting audiences' exposure to diverse information. The rationale for this goal has been that accurate beliefs and perceptions are necessary for good decision-making and for good governance in democratic society, and that development of these accurate beliefs requires some degree of exposure to information that challenges one's existing beliefs and opinions (Frey 1986, Hart et al 2009).

The Internet has brought more choice for what news and information individuals can access. Observers have warned that existing media policies are ill-suited for guaranteeing exposure to political diversity in the Internet age, as people are able to choose more freely from an ever-increasing variety of sources, many of which cater to and present a very narrow range of viewpoints. Given this range of choice, they argue, Americans will increasingly live in ideological echo chambers and polarization of different political groups will increase (Sunstein 2001, Prior 2007). Republicans and Democrats already read different newspapers and watch different TV news stations (Stroud 2007, Morris 2007, Iyengar & Hahn 2009). They read different political books (Krebs 2008). They even live in different places (Bishop 2008). The sources to which they attend are echo chambers: left-leaning and right-leaning blogs rarely link to each other (Adamic & Glance

2005) and comment threads contain little disagreement (Gilbert et al 2009). If people prefer to avoid hearing challenging views, we may see even greater political fragmentation in information consumption as people get better tools for filtering the news based on their own reactions and reactions of other people like them.

In my dissertation research, I sought to understand individuals' preferences for opinion diversity in the political news they access online (Chapter IV) and to develop selection (Chapters II-III) and presentation (Chapters IV-V) techniques that increase the amount of diversity in the political opinions people access through online news aggregators. I also studied whether there are other, nonpolitical spaces where exposure to diverse political views may already occur. I found that around half of the total political posts on *Blogger* occur on non-political blogs and I argue that understanding the political discussion in such spaces should be part of our community's research agenda (Chapter VI).

Before discussing this research, I will motivate my studies by briefly reviewing selective exposure theory, why exposure to diverse points of view is important, the evidence from online environments for and against the type of reading behavior selective exposure theory would predict in an environment of choice in information, and how these issues relate to news aggregators and other sources of political information online.

Selective exposure theory

Long before the Internet, researchers observed that people prefer to access information that suits their own opinions and worldview and to avoid information that challenges these beliefs (Lazarsfeld et al 1944, Klapper 1960, Berelson & Steiner 1964, Sweeney & Gruber 1984). This behavior is one way that people can avoid the uncomfortable feeling of cognitive dissonance (Festinger 1957). By selecting confirmatory information, they need not reevaluate their existing information or stance and they can reaffirm that they have correct beliefs (Kastenmüller et al 2010).

Selective exposure theory claims that people prefer to access supporting information and avoid challenging information. When people behave consistently with these preferences, the information they access supports their existing beliefs and preferences, and they are not prompted to go through the difficult process of reevaluating their views. Selective exposure literature also describes some conditions under which people may or may not act according to that preference (Frey 1986). For example, people are less likely to prefer a diet of only agreeable information when they are particularly curious about the topic (Frey 1986), when potentially discordant information is expected to be particularly useful – such as when they are skeptical of their own beliefs and seeking information rather than confirmation (Freedman 1965), or in response to a fairness norm (Sears 1965). The preferences described by selective exposure theory also predict how people will behave in information environments that offer choice between agreeable, disagreeable, and diverse information: given sufficient choice, they will read primarily agreeable information.

Dangers of selective exposure to political information

This prediction alarms many political theorists, who see exposure to diverse information as a prerequisite for many positive outcomes. Society risks losing these positive outcomes if people prefer to access only agreeable information and are able to construct a political information diet free of discordant information. In this section, I review three positive outcomes that can only occur if people are exposed to diverse political news, opinions, and information. First, exposure to diverse views is a necessary ingredient in deliberative debate, which political theorists argue is necessary for a healthy democracy. Second, counter-attitudinal information is necessary for people to learn and for better problem solving. Finally, understanding the distribution of opinions is necessary for people to accept the legitimacy of decisions with which they may not agree. I discuss these in more detail below.

First, exposure to diverse viewpoints is one of the necessary prerequisites for ideal deliberative debate, along with participants who are open to changing their minds and the formulation of arguments in terms of common interests rather than only in terms of competing interests of subgroups (Habermas 1962, Dewey 1954). Theorists argue that democracy flourishes in societies where political discussion is frequent and frequently approaches these deliberative ideals (Cohen 1989, Gutmann & Thompson 2004). If people only seek out and expose themselves to agreeable political news or discussion, this debate will not occur. Political theorists predict negative consequences resulting from its absence. Deliberation experiments have shown that interaction with like-minded people leads to polarization: participants tend to end up with more extreme views than they started with (Brown 1986, Sunstein 2002, Schkade et al 2007). Selective exposure to reinforcing news and opinion articles might also lead to opinion shifts to more extreme positions, and fragmentation of the audience to different spaces and sources may lead to discussion of articles that leads to even further polarization. Increased polarization would make it harder for society to find common ground on important issues.

Second, exposure to diverse opinions promotes learning and better problem solving. When people hold inaccurate beliefs, they must be confronted with information that challenges these beliefs in order to correct their understanding (Frey 1986, Hart et al 2009). Inclusion and consideration of diverse opinions leads to more divergent, out of the box thinking, which can improve individual and group problem solving and decision-making (Nemeth 1986, Nemeth & Rogers 1996). Through exposure to multiple viewpoints and perspectives, people become more aware of relevant information and are more able to think through all of the outcomes of a decision, and so societies will make better collective choices on important matters at all levels of government (Benhabib 1996).

Finally, there is a natural tendency for people, particularly those in the minority, to think that their own views are more broadly shared than they actually are (Sanders & Mullen 1982). This tendency is known as the false consensus effect: people can increase

their self-esteem by seeing their views as normative (Ross et al 1977). Having a better assessment of their true popularity may lead people to accept the legitimacy of disagreeable outcomes in the political sphere, rather than concocting conspiracy theories to explain how the supposed majority will was thwarted. If citizens feel they have been heard, and have heard others' reasons, they are more likely to accept a decision's legitimacy even if they do not get what they want (Benhabib 1996). Even when people reach different conclusions after hearing all of the sides, exposure to and consideration of different opinions persuade participants that the opponents' views have merit (Gutmann & Thompson 2004). Achieving this goal, however, can be difficult. When people receive evidence that their views are not normative, they can continue to increase their self-esteem by perceiving people with contrary views as acting according to a situational constraint rather than according to their own attitudes (fundamental attribution error, Jones & Harris 1967, Ross 1977) or as defective in some way (Ross et al 1977).

The dangers of selective exposure and the benefits of considering diverse and challenging points of view make it a reasonable public policy goal for people to be exposed to viewpoints other than their own. The principle of broad distribution of information from a diverse array of sources has long been a part of American media policy, going back to the First Amendment (Gentzkow & Shapiro 2008, Stucke & Grunes 2011). In *Democracy in America*, de Tocqueville wrote about his surprise at the number of early periodicals and their willingness to criticize those in power (volume 1, 1835). He also praised these periodicals, often quite partisan, for their role in bring people together and forming associations (volume 2, 1840). In the twentieth century, newspapers and broadcast news provided a digest of topics and editorial opinions for public discussion, and, through that common experience, serendipitous encounters with viewpoints other than their own (Sunstein 2001). The equal time provision and fairness doctrine, no longer in place, as well as antitrust policies toward media companies, have been designed to

ensure that these information sources present a range of viewpoints, within the same broadcast source or across different sources (Barron 1968, Stucke & Grunes 2011).

Sunstein and others, however, raise alarms that the Internet's increased choice of news sources and better tools for filtering out disagreeable news will undermine the role of media in presenting people with diverse viewpoints (Sunstein 2001, Pariser 2011). Whether the goal of exposure to diverse political views can be achieved in an environment of individual freedom, however, depends on whether designers succeed in selecting and presenting single collections that appeal to people with varying political views, or on whether they succeed in creating personalized collections that contain significant challenging information but are liked as much or more than collections without such challenging information.

Selective exposure and the Internet

Even when people prefer agreeable information, environments with limited numbers of sources can prevent the expression of such preferences. Research shows that people act on selective exposure preferences by selecting among sources rather than by selecting items from sources (Lowin 1967, Prior 2007). Though people might prefer a source with entirely agreeable items, if their only source choices all present some disagreeable information, they will not avoid reading altogether. Thus, in an environment where a modest number of media channels served a broad audience, mainstream channels could broadcast the news with enough diversity to appeal to a good-sized audience, without a risk of being avoided for that programming, and so their audience would have some exposure to informational programming that included challenging opinions. Policy decisions guaranteed a certain level of diversity among broadcast media, and a large portion of the news audience received these reasonably balanced broadcast sources simply by leaving the television on after a favorite program, the so-called inadvertent audience (Prior 2007,

Bennett & Iyengar 2008). Throughout the latter half of the twentieth century, this limited the effects of selective exposure preferences in determining individuals' actual exposure.

Technologies that offer more choices and personalization, however, undermine this role of broadcast media. They allow for the delivery of more channels, which have economic incentives to cater to niche audiences (Mullainathan & Shleifer 2005). Considerable evidence shows that people self-segregate into ideological television viewing groups (Pew 2004, Stroud 2007, Iyengar & Hahn 2009), while broadcast news audiences have declined precipitously (Bennett & Iyengar 2008). Time-shifting, watching programming on DVD or on-demand, and watching prime-time entertainment programming on channels without news shows has diminished the audience of people who inadvertently receive broadcast news by simply leaving on the television. Nearly half of Americans watched one of three broadcast news programs in 1970; this had dropped to ten percent by 2007 (Prior 2007). Those who seek out news, rather than merely leaving the television on, can and do choose news programming that caters to ideological niches (Morris 2005, Morris 2007). As people have greater choice in sources, they have greater potential to avoid disagreeable news sources or news sources altogether.

The Internet allows for even larger numbers of niche sources and for personalization of the sources that do have broad appeal. In *Being Digital* (1995), Negroponte celebrates the potential for a "Daily Me" – a highly personalized and tailored news source,[1] but for scholars concerned about negative societal consequences of political selective exposure (e.g. Sunstein 2001), the prospect of such news sources is alarming. Research presents conflicting evidence about the extent to which people prefer agreeable political information and are using the Internet act on those preferences: some research finds that people use the Internet to access mostly agreeable political information, while other studies suggest that people use the Internet to seek out a greater variety in sources

---

[1] Though Negroponte also allowed that *some* serendipity is of value, and proposed that people would read a "Daily Us" on a "rainy Sunday afternoon."

and to become more aware of political news and events. I describe each line of research in turn.

Less than two decades after Negroponte proposed the "Daily Me," substantial evidence supports the notion that people selectively expose themselves to agreeable political opinion in their choices of online information sources. Many surveys of individuals' preferences for news and studies of the traces of behavior on political information sites support the conclusion that people seek out agreeable political information to the exclusion of counter-attitudinal information. In a 2010 survey, 67% of Americans said that they only follow news or topics of interest to them, and 42% of American online news readers said that a key factor in their news site selection is the ability to customize the news the site presents them (Purcell et al 2010). Adamic and Glance found that political bloggers predominantly link to like-minded bloggers (2005), and Lawrence et al (2010) found substantial ideological segregation among political blog readers. Among the comments on political blogs, Gilbert et al found a 9:1 agreement to disagreement ratio, and that similar echo chambers exist on other blog genres as well (2009). During the 2004 US election cycle, people's political preferences motivated their choice of information sources across media types, including cable news and the Internet (Stroud 2007).

It is not clear, however, that everyone prefers to be exposed to only reinforcing viewpoints or that everyone is taking advantage of the Internet's affordances to act on these preferences. Sears and Friedman reviewed the literature from the 1950s and 1960s, finding five studies showing a preference for supportive information only, five showing a preference for diversity, and eight inconclusive (1967). In at least one face-to-face deliberation environment – structured, small group discussion of healthcare refrom in California – people fairly uniformly preferred groups that were moderately heterogeneous, neither homophilous nor highly discordant, in their viewpoints (Esterling et al, n.d.). Some recent studies support the idea that individuals are using the Internet to

seek out a broad range of political opinion and information, a behavior that we will call *diversity-seeking.* Stromer-Galley found that participants in online political discussions say that they want and seek out discussions with a diverse range of views (2003). In a summer 2004 survey, Horrigan et al found that Internet users were more aware of a diverse range of opinions than non-Internet users and that they were not using the Internet to filter out news and opinion that disagreed with their views (2004). Kelly et al found a diverse exchange of views within political USENET groups (2005): indeed, a good predictor of whether one person was generally liberal was whether the person's respondents were generally conservative in their posts. They did not see, however, much respect expressed between different positions on an issue (as the authors put it, "JS Mill would approve"). Papacharissi argues that the heated exchanges on news groups, for all their flaming, insults, and incivility, are a positive sign that different opinions are colliding and being contested (2004). Even if one favors Mills and Papacharissi's views of deliberative debate, Kelly et al's findings may not be a particular comfort, as USENET is rapidly falling out of favor.

Another view is that people seek out supportive information but are not averse to reading challenging information when they do encounter it. In an online experiment, Garrett found that subjects recruited from the readership of left-wing and right-wing sites were, on average, attracted to news stories that reinforced their viewpoint but showed only a mild aversion to clicking on stories that challenged them. Once they looked at those stories, however, they tended to spend more time reading them (Garrett 2009). Garrett proposes a theory, which I call *support-seeking*, that people seek out supporting items, but are indifferent about challenging items they encounter, so long as they see a sufficient number of supporting items. Gentzkow and Shapiro analyzed 2009 traffic data from comScore's panel of over one million US Internet users and found that ideological isolation in people's online news sources was lower than many feared, and lower than their

day-to-day face-to-face interactions, yet higher than their ideological isolation for most offline news sources (2011).

The mixed, and apparently contradictory, results leave two important unanswered questions. To what extent do people prefer to access agreeable political information while avoiding challenging material? To what extent, and how, are they are using filtering and niche-appeal Internet sources to act on such preferences?

News aggregators

There are many websites that aggregate political news and opinion, such as *Digg*, *Reddit*, *Google News*, and *Memeorandum;* these are one type of online space where selective exposure preferences could be particularly consequential. People, particularly younger individuals, are increasingly relying on these sites for their news access. In 2010, editorially and algorithmically curated news aggregators (e.g., *Google News*) were the most popular news source for online news users (used by 56% of adult Americans – 68% among the 18-29 demographic – who access news online on a typical day), while 7% accessed reader-curated news aggregators (e.g., *Digg*) on a typical day (Purcell et al 2010). Both types of aggregators present links to recent articles from blogs and commercial media, and some of them also offer a forum for readers to discuss the linked stories. As the news aggregator marketplace matures, consumers will have many options to choose from and, over time, they will likely gravitate to aggregators that offer the mix of articles and the type of discussion that they like best, though other factors, such as price and usability, will also influence their choice.

Selective exposure theory predicts a preference for news aggregators that provide readers with only agreeable sets of items. As reviewed previously, though, studies of online political information diets show mixed results, so it remains an open question what mix of agreeable and challenging articles people would ideally like to see on these sites. As online news aggregators, and possibly personalized news aggregators, become an important

destination for Americans' news, scholars and policymakers concerned about selective exposure preferences, behavior, and outcomes will want to know the answer to this question. If people do prefer agreeable political information, and actively seek it out to the exclusion of challenging information, a new question emerges: are there techniques for selecting and presenting political news that can encourage people to seek out or at least tolerate sources or items that present diverse views?

The designers of news aggregator sites have a more immediate interest in knowing what mix of agreeable and challenging items people want to see, as they want to deliver collections of items that make readers want to come back to their sites. From a designer's perspective, there are actually two versions of this question, depending on whether the news aggregator will present each reader a potentially different collection of articles, or whether the same collection will be presented to everyone. If personalized collections will be presented, the question is, "what is the optimal percentage of agreeable items to present?" If the same collection will be presented to everyone, the question is, "is it possible to keep a set of readers with diverse political preferences satisfied or will groups with different political preferences inevitably drift towards using separate aggregator sites?" The latter is a very real possibility, as the launch of several avowedly conservative competitors to *Digg* (e.g. *Lively Links*, *GOPHub*, and the somewhat awkwardly named *R-igg)* or *YouTube* (e.g. *PopModal.com*) might suggest, though none of these have yet gained the popularity of *Digg* or *YouTube,* and many have gone defunct.

Non-political spaces

Though policy has focused on exposure to diverse information in news and other formal spaces for political discussion, these are not the spaces that political theorists have most celebrated and cherished for exposing people to diverse news and opinion. Habermas talks about a "public sphere" that brings together people with diverse backgrounds and grants them a certain amount of equality as discussants (1962). Eighteenth century growth

in journals, lodges, salons, and coffeehouses nurtured a space to talk about politics outside of the control of the state. In these discursive arenas, individuals' status was largely disregarded, many had access and were included, and participants would raise issues of public concern for discussion. Putnam, in his studies of democracy in Italy (1993) and later social capital in America (2000), also celebrated the role of social clubs and activities that bring people together for non-political reasons. In choral societies, bowling leagues, Kiwanis clubs, and parent teacher associations, participants would develop relationships around the shared activity, but this social capital could be repurposed as ideas and information, including diverse viewpoints, flowed through these ties. Sunstein expands important non-political spaces to include those that promote less formal or protracted interactions: the parks, the streets, and the corner market (2001). He describes these spaces as "glue" holding diverse groups together. People who may otherwise encounter primarily views that affirm their own can have serendipitous encounters in these shared spaces.

Non-political, public spaces many not only bring together people with diverse views and backgrounds, but could also do so in a way that has further benefits. Members' desire to protect their social ties within these spaces provides participants with an incentive to keep the discussion polite (within the norms of the relationship), to phrase arguments in a broader perspective, and to listen to each other (Putnam 2000). Unfortunately, the desire to protect their existing relationships in these spaces may suppress the expression of and exposure to challenging views. Americans are much less likely to bring up politics in social contexts with diverse views (Noelle-Neumann 1993). In more recent work, Mutz raises doubts about non-political settings' abilities to trigger and nurture cross-cutting debate (2006). Many voluntary social ties, such as golf club memberships, are formed among people from similar backgrounds who share political preferences. Mutz actually found a negative correlation between an individual's number of voluntary social associations and their exposure to cross-cutting debate. Among voluntary

associations, parent teacher organizations appear to be an exception to this rule, and some workplaces also expose individuals to a broader range of opinions (Krebs 2005, Mutz & Mondak 2006). Even when Americans do voice political opinions, though, they often go to great lengths to distance their opinion from politics (Eliasoph 1998). They preface their opinions with statements like "Not to be political but…" and try to cast issues as local or personal arguments rather than as part of a larger political debate.

On the Internet, there are many spaces, such as hobby forums, various interest-based or personal blogs, and gaming sites, where people come together for shared interests. Might these sites be important spaces for cross-cutting political debate? If so, could they provide some of the serendipitous exposure to other viewpoints that Sunstein and others worry we are losing on the Internet? Or might the Internet's affordances for self-censorship, moderation, and channel-switching mean that people go elsewhere to bring up potentially contentious issues? Non-political spaces online have not historically received much attention from researchers studying politics online, but it is important to understand if political discussion is occurring in them, and, if so, the characteristics of that discussion.

Outline of the dissertation

A broad research agenda is required to understand individuals' preferences for political information preferences online, and ways to nudge preferences or behaviors. With a focus on news and information sites, researchers must work to (1) understand existing preferences and behaviors for accessing political information online, (2) explore ways to nudge challenge-averse individuals to read or consider more diverse political news, (3) study techniques to help people engage with diverse points of views, and communicate their own, and (4) evaluate the effects of promoting a diverse information diet. Outside of news sites, researchers should study non-political spaces to identify whether political discussion occurs there or if politics is treated as taboo. If political discussion does occur

in non-political spaces online, researchers should examine whether it offers visitors to these sites serendipitous exposure to diverse political views.

This research agenda is broader than any one dissertation. In my work, I decided to take on elements of the first two challenges – understanding people's preferences and behavior and developing and evaluating ways to select and present more diverse content – and the fifth challenge, measuring and examining the political discussion that occurs in one type of non-political space. Table I-1 presents a summary of my questions and findings, by chapter. I did not study ways to increase engagement or the consequences of exposure to diverse political views on idea formation, civic engagement, or political participation. For the purposes of this work, I limited my definition of diversity to collections representing multiple views in order to achieve the goal of exposing people to information that may be discordant with their current beliefs. There are other facets of diversity that likely matter (for a good discussion, see Garrett & Resnick 2011), but they are beyond the scope of this research.

In studies of news aggregators, I found that people as a species neither inherently avoid challenging information nor inherently seek out diversity, but that there are individual differences (Chapter IV). I also explored ways to select diverse content for news aggregators and to present that content to challenge-averse individuals. In Chapter II, I present the *Sidelines* algorithm, which selects more representative content from user votes (such as those used on news aggregators such as *Digg* and *Reddit)*, and, in Chapter III, an evaluation of this algorithm. I also tested presentation techniques intended to make challenging opinions more appealing to people who would prefer not to see them and I outline a design space and some additional examples of presentation techniques that should be evaluated in future research (Chapters IV-V).

Inspired by political theorists' arguments that non-political spaces have historically been an important site for exposure to cross-cutting ideas and discourse, I investigated non-political blogs to see to what extent they contribute to political discourse online. I

found that a substantial amount of political discussion online occurs in non-political spaces: on *Blogger*, about half of the political discussion occurs on non-political blogs, and when it occurs, it is not treated as taboo (Chapter VI). These non-political spaces may be different than the more frequently studied online political spaces, and better understanding the idea formation and discussion that happens in them will be an important future research topic.

I conclude by summarizing what we learned from these studies, reviewing other recent advances on the topic of selective exposure and online political discussion, and outlining next steps in this research area.

Topic		Ch	Question	Findings and outcomes
	Diversity preferences	IV	• With what percent of agreeable and disagreeable items are people most happy?	• Some people seek diversity, while others avoid challenge.
News aggregators	Selecting diverse collections	II	• What are some measurable diversity goals for a selection algorithm? • Can we design an algorithm that produces more diverse collections of items than a pure-popularity algorithm?	• Specification of inclusion, alienation, and representation measures. • Development and assessment of the Sidelines algorithm on *Digg* and blog data.
		III	• Do people notice the difference in diversity between collections generated by Sidelines and by Pure Popularity?	• Yes, they do.
	Presenting diverse collections	IV	• Does highlighting agreeable items increase satisfaction with collections that include challenging items? • Does highlighting agreeable items and placing them first increase satisfaction with collections that include challenging items?	• Not overall. Highlighting agreeable items made peoples' reactions stronger. • No. Highlighting agreeable items and placing them first made people less satisfied overall.
		V	• What is the design space for visualizations intended to increase the balance of political items that people read online? • How can / should options within this design space be evaluated?	• A sketch of a design space, including what information can be presented and when. • One completed but flawed experiment, lessons learned, and a new design.
Non-political spaces		VII	• Does political discussion occur in non-political spaces online? • If so, do people treat it as taboo?	• On *Blogger*, half of the political discussion occurs on non-political blogs. • When politics comes up on non-political blogs, it elicits as many comments as non-political posts. These comments engage the political content and include some disagreement.

Table I-1 Summary of questions and findings in this dissertation.

CHAPTER II

Sidelines: An Algorithm for Increasing Diversity in Voter-curated News Aggregators[2]

Online news and opinion aggregators have become an increasingly important source of political information. Some sites, such as *Digg* and *Reddit*, rely on reader votes to select news articles and blog entries to appear on their front pages. They have become popular — *Digg*, for example, gets more than 35 million visitors each month. *Memeorandum,* a similar site, selects political news articles and blog entries based in large part on the links among stories: those articles with more incoming links from more popular sources are more likely to be selected. Many aggregators also convene conversations around the articles selected for the front page. Since these sites serve as an important source of political information, site designers might have an interest in ensuring that the items listed represent a diversity of viewpoints. But before these sites can highlight diverse stories, they need to be capable of measuring the diversity of their collections and selecting stories that represent varied viewpoints.

Even if a site selects items based on votes or links from people with diverse views, algorithms based solely on popularity may lead to a tyranny of the majority that effectively

---

[2] Portions of this chapter, and the next, first appeared as Munson, SA; Zhou, DX; Resnick, P. 2009. "Sidelines: An Algorithm for Increasing Diversity in News and Opinion Aggregators," *ICWSM 2009.* Daniel Zhou conducted the *Digg.com* evaluation. Anton Kast at Digg framed the problem as one of adding diversity heuristics to a pure popularity algorithm.

suppresses minority viewpoints. That is, even if there is only a slight bias on the input side, there can be a large bias on the output side, a tipping toward the majority viewpoint. For example, if a site has 60 left-leaning voters and 40 right-leaning voters, and each can vote for many articles, then it may be that all the left-leaning articles will get more votes than all the right-leaning articles. Similarly, if a link-following algorithm such as PageRank (Brin & Page 1998) is used on a corpus of blog posts that has 60% left-leaning authors, the left-leaning posts could easily make up 100% of the top ranking articles. If a news aggregator takes no corrective steps, the minority may feel disenfranchised and abandon use of the site. This could happen even if they would have been happy to stay and participate in a site that included their viewpoint only 40% of the time. Over time, even people who would prefer to get a mixed selection of news, and to participate in discussions with people who have mixed views, would end up sorting themselves into homogeneous groups.

Topic		Question	Findings and outcomes
News aggregators	Selecting diverse collections	• What are some measurable diversity goals for a selection algorithm?	• Specification of inclusion, alienation, and representation measures.
		• Can we design an algorithm that produces more diverse collections of items than a pure-popularity algorithm?	• Development and assessment of the Sidelines algorithm on *Digg* and blog data.

Table II-1 Summary of questions and contributions in this chapter.

Diversity goals

Beyond retaining readers with minority viewpoints, there are several societal reasons why some form of diversity might be valuable. One diversity goal is to make as many people as possible feel that their viewpoint is represented in the aggregator's result set. Groups and society as a whole make better decisions when they have diverse participants and those participants' viewpoints are all taken into account (Hong & Page 2001). People who feel that their view is a minority position and so far unspoken may remain silent to promote social harmony (Rossenburg 1955; Mansbridge 1980): by making people see that their

viewpoints are publicly represented in the selected news and opinion items, people may be more likely to articulate their viewpoints in discussion, at the news aggregator site and elsewhere. Moreover, people may be more open to hearing challenging opinions once they feel their own viewpoint is represented (Garrett 2009), so making more people feel included may induce more people to expose themselves to challenging viewpoints.

A second diversity goal is to represent viewpoints in the result set in proportion to their popularity. This could help everyone to understand the relative popularity of different viewpoints. People, particularly those in the minority, tend to believe that their own views are more broadly shared than they actually are (Ross et al 1977; Sanders & Mullen 1982). Having a better assessment of positions' true popularity may lead people to accept the legitimacy of disagreeable outcomes in the political sphere, rather than concocting conspiracy theories to explain how the supposed majority will was thwarted.

A third diversity goal is to ensure that everyone is exposed to challenging viewpoints. A long history of experiments has shown that deliberation on an issue with like-minded people leads to polarization: everyone tends to end with more extreme views than they started with (Brown 1986, Schkade 2007). Awareness of minority views can also lead individuals in the majority to more divergent, out of the box thinking, which can be useful in problem solving (Nemeth & Rogers 1996).

It will be valuable, then, to develop algorithms for news and opinion aggregators that select items that in some way reflect the diversity of opinions of their readers.

Overview of the chapter

In this chapter, I present an algorithm, which I call Sidelines, that is intended to increase the diversity of result sets. I compare the results produced by the algorithm to those from a Pure Popularity selection algorithm using three diversity metrics. In the following chapter, I report on an online experiment that asked people to assess the result sets subjectively.

Data sets

The first domain we explored consisted of user votes on *Digg*.com. Using *Digg*'s public API, we tracked items and votes in the category "World and Business", which includes political news and opinion, between 11 October and 30 November 2008. This category had an average of 4,600 new incoming stories and 85,000 diggs (votes from users to stories) from an average of 24,000 users every day. Voting roughly followed a power law – 91% of users voted less than once per day, contributing 28% of the total votes, and 0.7% of users voted more than 10 times per day, contributing 32% of the total.

The second domain consisted of links from blog posts from a collection of 500 political blogs. We selected blogs for this panel from the *Wonkosphere* directory of political blogs (1,316 blogs). To be included, a blog had to publish the full content of its posts, including markup, as an RSS or Atom feed, to have posted a blog entry within the previous month, and to have most of its front-page posts be about political topics. This left us with less than our goal of 500 blogs, so we selected others for inclusion by examining the link rolls of blogs already in the sample, until we reached 500.

We coded each of the source blogs based on its political ideology (liberal, independent, or conservative). We consulted both *Wonkosphere* and *PresidentialWatch08*, which maintain directories of weblogs classified by political affiliation. In addition, one of the authors read entries from each blog and coded it manually. When the three classifications disagreed, the majority classification prevailed. If a blog was only classified by one of *Wonkosphere* and *PresidentialWatch*, and there was disagreement between that source and the reader, we chose the blogger's self-identification (if present) or the third-party (*Wonkosphere* or *PresidentialWatch*) assessment. Our panel of blogs contained 257 liberal blogs (52%), 174 conservative blogs (35%), and 63 independent blogs (13%).[3]

---

[3] Six additional blogs that we tracked did not post entries during the period October 26 to November 25, 2008, and are omitted from this classification. A full list of sources and classifications is available at http://www.smunson.com/bloggregator/sources-icwsm.csv.

For use in our selection algorithms, we think of blogs as voters, links as votes, and the web pages that they link to as the candidate items for selection. Note that blog entries often link to news stories from mainstream media sites as well as other blog entries and non-blog web pages. Thus, the universe of items was not limited to the entries in the set of 500 blogs. To avoid bootstrapping issues, we collected links to items for two days before the time period during which we generated results. During the period 24 October to 25 November, there were a total of 166,503 links to 106,765 distinct items.
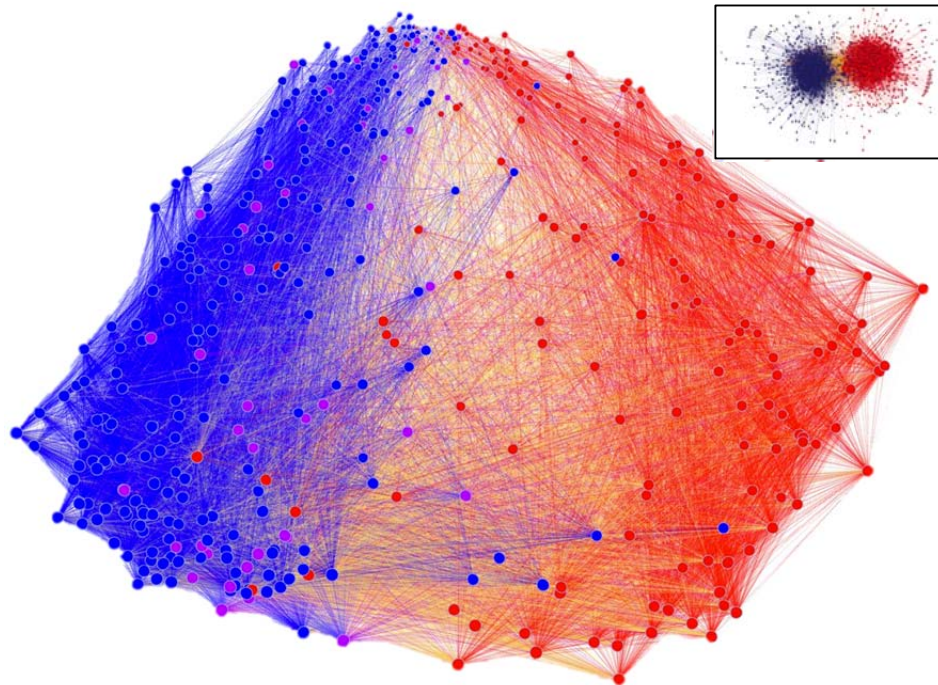
Figure II-1 Visualization of the projection of bipartite graph of blog-item links. Multidimensional scaling layout according to Jaccard similarity using GUESS. The link patterns are polarized, much like the behavior observed by Adamic and Glance between blogs (inset). Nodes are sized according to degree; edges indicate Jaccard similarity above average with edge width according to the Jaccard similarity. Link colors match the node colors when the blogs being linked have the same color; maroon connects independent and conservative, violet blue connects liberal and independent, and yellow connects liberal and conservative**.**

Liberal and conservative blogs tended to link to different items. To visualize this, we created a graph with the blogs as nodes and a link between two blogs if the two voted for (i.e., linked to) at least six items in common (Figure II-1). Extrapolating from the color-coding used by Adamic and Glance (2005, Figure II-1 inset), we colored the nodes blue (liberal), red (conservative), and purple (independent).

Algorithms

We implemented two algorithms. The first, a generalization of approval voting, served as a comparison point for the second, which was intended to increase diversity. In an approval voting system, each voter votes for as many candidates as desired and the top k vote-getters are selected (Brams & Fishburn 1978). With the blog collections, we treated each blog as a voter and a link from a blog post to an item as a vote from the blogger for the item. Because reader interest in news decays with age (hence the term news), we modified the approval voting system to decay each vote (digg or blog link) linearly over time, such that:

$$w = \begin{cases} 1 - t/t_{max} & \text{for } t < t_{max} \\ 0 & \text{for } t \geq t_{max} \end{cases}$$

where w is the vote's weight, t is the time since the link was first detected, and t_{max} is the time after which the vote no longer contributes to the total. The popularity of an item was computed as the sum of the time-weighted votes, Σw. The algorithm then selected the top k items in terms of total time-weighted votes (Algorithm 1). We call this the *Pure Popularity* algorithm.

The second algorithm, which we call the *Sidelines* algorithm, dynamically suppresses the influence of voters (Algorithm 2). As with the pure popularity algorithm, it incrementally selects the item with the highest total time-weighted votes at each step. After selecting an item, however, anyone who had voted for it is relegated to the sidelines for a few turns; that is, votes from sidelined voters were removed from consideration for the selection of the next few items. Note that the Sidelines algorithm does not take into account the group affiliations of voters or correlations between them. Our goal in this study was to investigate whether simply reducing the influence of voters whose preferred items had already been selected would improve the diversity of the result sets.

Algorithm 1 Pure Popularity

Param: I {The set of items}
Param: U {The set of users}
Param: V {The users by items matrix of votes; V_{ui} = 1 if user u voted for item
  i, else 0}
Param: k {Number of items to return}
Param: timeweight() {A function that time-weights votes based on age}

results ← []

for all items i **in** I **do:**
  $i.score$ ← 0
  for all users u **in** U **do:** $i.score$ ← $i.score$ + timeweight($V[u,i]$)

results ← first k items in I sorted by score

Return: *results*

Algorithm 2 Sidelines

Param: I {The set of items}
Param: U {The set of users}
Param: V {The users by items matrix of votes; V_{ui} = 1 if user u voted for item
  i, else 0}
Param: k {Number of items to return}
Param: timeweight() {A function that time-weights votes based on age}
Param: *turns* {Number of turns to sideline a "successful" voter}

results ← []

for all users u **in** U **do:**
  sidelined[u] ← 0 {Initially no one is sidelined for any turns}

while length(*results*) < k **and** length(*results*) < length(I) **do:**
  for all items i **in** I **do:**
    $i.score$ ← 0
  for all users u **in** U **do:**
    if *sidelined[u]* ≤ 0 **then** $i.score$ ← $i.score$ + timeweight($V[u,i]$)

  {Decrease the sideline turns for each item}
  for all users u **in** U **do:**
    sidelined[u] ← *sidelined[u]* – 1

  winner ← item with maximum score
  Remove *winner* from I
  Append *winner* to *results*
  for all voters v **in** V **do:**
    if $V[u,$ *winner]* $= 1$ **then** *sidelined[u]* ← *turns*

Return: *results*

Diversity measures

I next review three metrics for diversity that reflect different diversity goals: inclusion, reducing alienation, and proportional representation.

Inclusion / Exclusion

One simple diversity metric measures the proportion of voters who had a voted-for item in the result; this is the inclusion score. When computing this metric for any snapshot of results, only voters who had voted for at least one item in the previous 48 hours are included, since votes decay over 48 hours in our time-weighting. The percent of voters who did not have any voted-for items in the result set is the exclusion score. A higher inclusion (and hence lower exclusion) score is one indicator of greater diversity.

Note that we would expect the Sidelines algorithm to include items for at least as many voters as pure popularity, since it gives more weight to votes from voters who do not yet have an item included. Because it is a greedy algorithm that selects the most popular item at each step, pathological cases exist where the Sidelines algorithm actually reduces inclusion.

Alienation

A more sophisticated version of exclusion measures the position of the best item in the algorithm's result set rather than just whether any voted-for item is present. The measure generalizes from the Chamberlin-Courant scoring rule for voting systems that select a set of candidates for a committee (1983). The ideal committee is one that minimizes the total alienation of the voters, measured as the sum of all the voter's alienation scores. Finding an ideal committee according to this sum of alienation scores has been shown to be NP-Complete (Procaccia et al 2008).

In our case, we have a sparse set of approval votes from each voter rather than a complete ranking. Moreover, it is natural to think of the result set $K=\{k_1, k_2, \ldots, k_{|K|}\}$ as

ordered, since readers will notice the top news stories in a listing before they notice ones lower down. We define the alienation score for user u against K as

$$S_{\text{alienation}}(K, u) = \begin{cases} \min(i) & \text{where } k_i \in K \\ |K| + 1 & \text{otherwise} \end{cases}$$

That is, $S_{\text{alienation}}(K,u)$ is either the position of the highest item in K that u voted for, or $|K|+1$ if K has no item that u voted for. We then define the overall alienation score for K as the sum of individual alienations, normalized by the maximum possible alienation so that values always lie in the interval $(1/(|K|+1), 1]$.

$$S_{\text{alienation}}(K) = \frac{\sum_{u \in U} S_{\text{alienation}}(K, u)}{(|K| + 1)|U|}$$

A lower score indicates improvement on this diversity metric: more people's viewpoints are represented higher up in the result set. As with the simple inclusion/exclusion metric, the Sidelines algorithm will normally decrease the alienation score at least modestly, though pathological cases exist where it could increase alienation.

Proportional representation

A third diversity metric is a generalized notion of proportional representation: we define a divergence scoring function that is minimized when the result set K has votes from different groups in proportion to their representation in the voter population. In the introduction and description of our pilot data, we divided the people and items into Red, Blue, and Purple, with Red people generally voting for or linking to Red items. If the user population were 60% Blue, we suggested that it would be better to select 60% Blue items than to select 100% Blue items, as might occur if we simply take the approval voting outcome.

More generally, suppose that there are groups, $G=(g_1, \ldots, g_{|G|})$, and that each person u may have partial affiliation with each group, which we represent by a vector $\mathbf{u}_G = (u_{g_1}, \ldots u_{g_{|G|}})$, with $\sum_{g \in G} u_g = 1$. For a set of users U, we define the representation for the groups as

$$U_G = \frac{\sum_{u \in U} \mathbf{u}_G}{|U|}$$

Note that, by construction, $\sum_{g \in G} U_g = 1$, $\Sigma u_G = 1$. That is, the weights express the proportion of total affiliation for each group. In the case where individual affiliations are pure (i.e, each person is affiliated with just one group), the representation vector simply expresses the proportion of users in each of the groups.

Given a set of votes V, for any item i we define i's representativeness with respect to the groups' preferences as a vector of weights $\mathbf{i}_G$, with

$$i_g = \frac{\sum_{u \in U} u_g v_{ui}}{\sum_{u \in U} v_{ui}}$$

That is, each vote is weighted by the portion of the voter's affiliation that belongs to the group. The sum of weighted votes divided by the total votes gives the proportion of the total votes that are affiliated with the group. Note that $\sum_{g \in G} i_g = 1$. In the case where individual affiliations are pure, i_g simply expresses the proportion of all votes for the item that came from users in group g.

Then, we define the representativeness vector $\mathbf{K}_G$ on a subset of items K as the mean representativeness over items in the subset:

$$\mathbf{K}_G = \frac{\sum_{i \in K} \mathbf{i}_G}{|K|}$$

Next we compare the two vectors $\mathbf{U}_G$ and $\mathbf{K}_G$ to compute the amount that the groups' preferences for the subset K diverge from the groups' proportional representation. Interpreting $\mathbf{U}_G$ and $\mathbf{K}_G$ as probability distributions over the groups, we compute the Kullback-Liebler divergence (1951), otherwise known as the conditional entropy:

$$D(\mathbf{U}_G \| \mathbf{K}_G) = \sum_{g \in G} U_g \log \frac{U_g}{K_g}$$

The divergence score is always positive. Lower scores indicate more proportional representation: the mean item representation score is closer to the mean of the user affiliation scores.

Digg.com evaluation

For the *Digg* data set, we computed result sets of size 35 for the pure popularity and Sidelines algorithms once per day from 19-30 November 2008. For the Sidelines algorithm, the turns parameter was set to 20, meaning that a voter's votes were excluded for the selection of the next 20 items after a voted-for item was selected. For both the Sidelines and Pure Popularity algorithms, each link counted as one vote when it was first detected and then decayed linearly to 0 over 48 hours (t_{max}).

Averaging over the 12 result snapshots, 65.1% of users who voted for at least one item in the previous 48 hours had at least one voted-for item included in the 35 results. For the Sidelines algorithm, an average of 66.8% of voters had at least one voted-for item selected. The difference is statistically significant (paired *t*-test, $t(11)=6.05$, $p<0.001$). Though statistically significant, this effect is quite small in practice: less than two additional users in one hundred would have found an item they voted for in the result set.

The alienation score was also lower (mean 0.476 vs. 0.463). Partly, this results from including a voted-for item for more users. In addition, contingent on having an item selected, voted-for items appeared somewhat earlier in the result sets: the mean position was 6.91 for the Sidelines algorithm and 7.12 for Pure Popularity ($t(179,668)= 5.63$, $p<0.001$).

We do not have a classification of *Digg* users into opinion groups. Therefore, we were not able to compute a divergence score to measure the proportional representation with respect to opinion groups of selected items vs. the overall population of voters.

Blog links evaluation

For the set of 500 blogs, we generated result sets of $k=12$ items using each of the algorithms. In the Sidelines algorithm, a blog sat out the voting for the next *turns*=7 items after an item it linked to was selected. For both the Sidelines and Pure Popularity

algorithms, each link counted as one vote when it was first detected and then decayed linearly to 0 over 48 hours (t_{max}). We generated results snapshots at 6-hour intervals for a period from 26 October to 25 November 2008.

The Sidelines algorithm achieved a somewhat higher mean inclusion score (0.445) than the Pure Popularity algorithm (0.419). The difference was statistically significant (paired t-test, $t(120) = 8.701$, $p < 0.001$). We also computed the alienation score, $S_{alienation}$ for each snapshot (Figure II-2). For this calculation, we included only the blogs that had linked to an item in the time window used to generate the snapshot. The mean $S_{alienation}$ for Sidelines result sets was 0.809, and the mean $S_{alienation}$ for Pure Popularity was 0.796. This difference is statistically significant (paired t-test, $t(120) = 7.864$, $p < 0.001$).
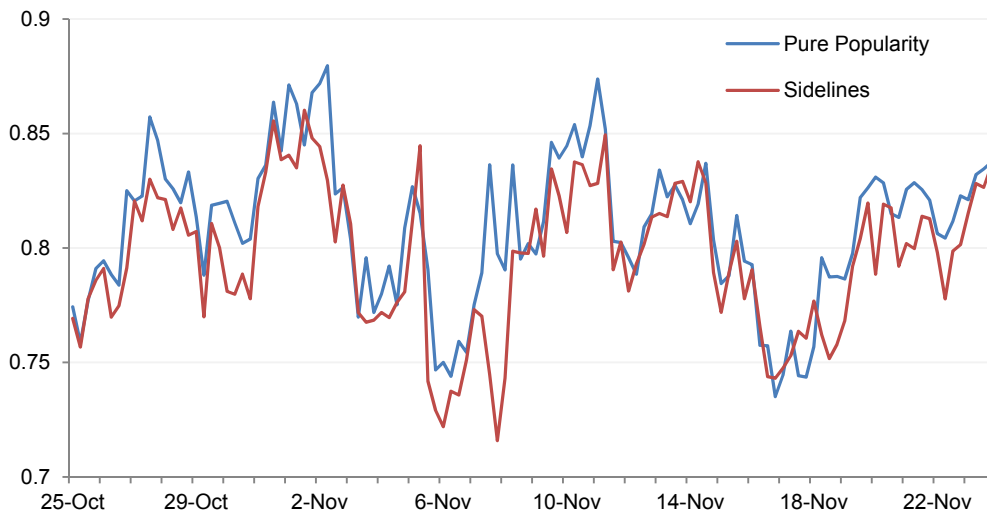
Figure II-2 Alienation score for Sidelines and Pure Popularity algorithms on blog data set.

As described previously, we expected that the Pure Popularity algorithm might tip toward producing very liberally biased results given that the sample of blogs had somewhat more liberal than conservative blogs, and we expected that the Sidelines algorithm would tip less. To evaluate this, we calculated the proportional representation divergence score, $D(\mathbf{U}_G||\mathbf{K}_G)$, for each snapshot (Figure II-3).
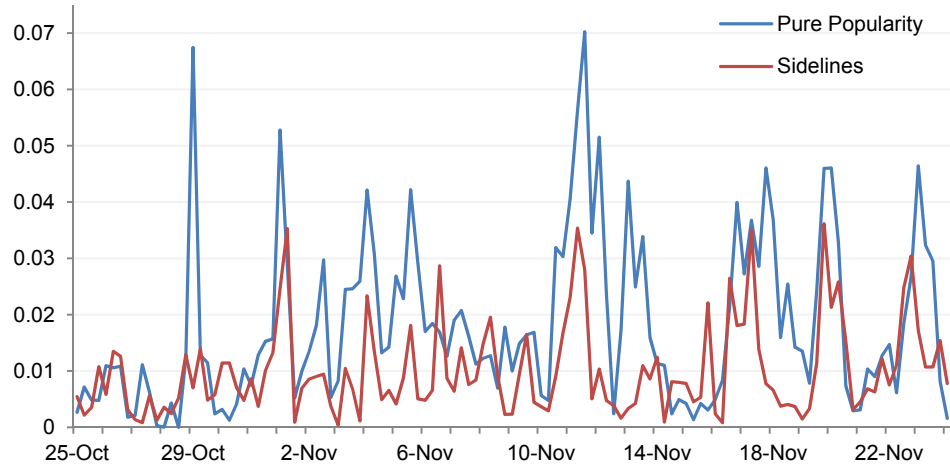
Figure II-3. Proportional Representativeness score for Sidelines and Pure Popularity algorithms on blog data set.

The Pure Popularity algorithm showed some evidence of the expected tipping. While 52% of blogs were classified as liberal (blue), the mean representation of blue opinion among the items selected by the Pure Popularity algorithm was 61.9% (See $\mathbf{K}_B$ column in Table II-2); the mean divergence score, $D(\mathbf{U}_G\| \mathbf{K}_G)$, which takes into account representation of blue, red, and purple, was 0.018. The Sidelines algorithm showed evidence of tipping as well, but not as severely; the mean representation of blue opinion was 58.6% and the mean divergence score was 0.010. The difference in divergences between the two algorithms was statistically significant (paired t-test, $t(120) = 6.953$, $p < 0.001$). Table II-2 summarizes the mean distributions of blue, red, and purple representation for each algorithm.

	$\mathbf{U}_B$	$\mathbf{U}_R$	$\mathbf{U}_P$	
Blog population	0.520	0.352	0.128	
	$\mathbf{K}_B$	$\mathbf{K}_R$	$\mathbf{K}_P$	$\mathbf{div}_{KL}$
Pure Popularity	0.619	0.277	0.103	0.018
Sidelines	0.586	0.313	0.101	0.010

Table II-2 Proportional Representation by algorithm.

Discussion

While the Sidelines algorithm improved the diversity metrics as compared to Pure Popularity, effects were not large. Especially with the *Digg* data, this may be due in part to the large number of voters. For example, in the snapshot taken for day November 22, there were 4476 votes for the most popular item. 670 of the 1769 voters for the second most popular item had also voted for the most popular item, but that still left 1099 votes for the second most popular item. We speculate that with so many voters on *Digg*, even when all the voters for the most popular item are removed, there are many like-minded voters left to vote for the next most popular item. Thus, among the twelve *Digg* snapshots, the median position in the result set where the Pure Popularity and Sidelines algorithm first disagreed was not until the fourth item. More importantly, the median position where an item first appeared in the top-35 results with Sidelines that did not appear at all in the Pure Popularity result set was position 19, with a range of 16-31.

There are two obvious weaknesses in the initial Sidelines algorithm that we used in the pilot work. First, it is more suited to maximizing the number of people who feel their viewpoints are represented in the result set than it is to achieving proportional representation of the different viewpoints of the population, two alternative notions of diversity. Second, because only voters who actually voted for an item sit on the sidelines, and not other people who share their viewpoint, it is not as effective as it could be at getting more viewpoints represented in the final set. Future work should explore ideas that might improve on the Sidelines algorithm, at least for achieving some types of diversity.

The first approach is to make adjustments to the Sidelines algorithm, changing who is sidelined and for how long. As shown in the pilot study, when we increased the number of rounds users were sidelined before their votes counted again, there was a greater change in the results list for *Digg*, as compared to the list generated from Pure Popularity voting. Future work could vary this number and find the optimal value, which

might depend on other parameters such as the size of the set of items, the size of the desired result set, the number of users, and the distribution of votes. There are a number of modest adjustments that could be made to the Sidelines algorithm, leading to a generalized form (Algorithm 3). This generalized algorithm includes the possibility for a gradual recovery from being sidelined, as well as boost phase for voters who have gone an excessively long time without having an item selected (Figure II-4).
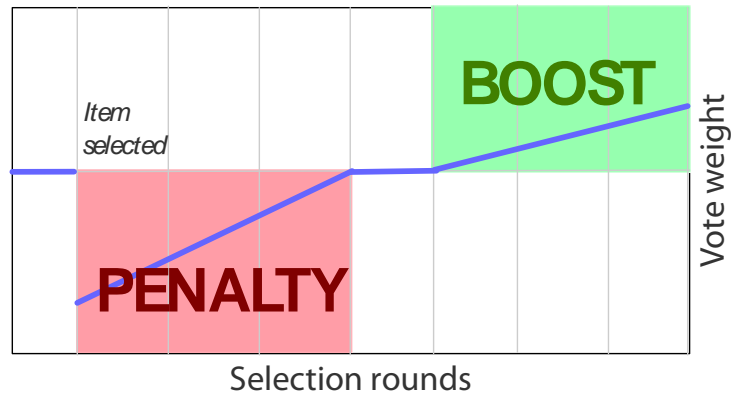
Figure II-4. Representation of the Generalized Sidelines algorithm.

Algorithm 3 Generalized Sidelines

```
Param: I {The set of items}
Param: U {The set of users}
Param: V {The users by items matrix of votes; Vui = 1 if user u voted for item
  i, else 0}
Param: k {Number of items to return}
Param: timeweight() {A function that time-weights votes based on age}
Param: PenaltyTurns {Number of turns to sideline a "successful" user}
Param: Recovery {1 if voters gradually recover from being Sidelined, else 0}
Param: Penalty {The "cost" to a user's votes for being Sidelined}
Param: Boost {1 increase weight of votes from users who have not had an item
  selected recently, else 0}
Param: BoostAfterTurns {number of turns before a user's vote weight starts
  increasing}
Param: BoostPerTurn {amount to increase a vote's weight during boost phase}

results ← []

for all users u in U do:
  SidelineCounter[u] ← 0 {Initialize to 0 for all users}

while length(results) < k and length(results) < length(I) do:
  for all items i in I do:
    i.score ← 0
  for all users u in U do:
    if SidelineCounter[u] < 0:
      {Not Sidelined}
      if (Boost) and SidelineCounter[u] ≤ -BoostAfterTurns
        {In the Boost Phase}
        i.score += timeweight(V[u,i]•(1+BoostPerTurn•(-SidelineCounter[u])))
      else:
        {In the Neutral Phase}
        c += timeweight(V[u,i])
    else:
      {Sidelined}
      i.score += timeweight(V[u,i]-Penalty+Recovery•(Penalty/PenaltyTurns))

  {Increase the turns since selection}
  for all users u in U do:
    SidelineCounter[u] += -1

  winner ← item with maximum score
  Remove winner from I
  Append winner to results
  for all voters v in V do:
    if V[u, winner] = 1 then SidelineCounter[u] ← PenaltyTurns

Return: results
```

To see how Pure Popularity, Sidelines, and other variations can be expressed in this generalized algorithm, see Table II-3.

	Penalty Phase			Boost Phase		
	Penalty Turns	Recovery	Penalty	Boost	Boost After Turns	Boost Per Turn
Pure Popularity	0	-	-	0	-	-
Sidelines	7	0	1	0	-	-
Sidelines with Recovery	7	1	1	0	-	-
Anti-Sidelines	0	-	-	1	7	0.1
Sidelines with Boost	7	0	1	1	7	0.1

Table II-3 Examples of Generalized Sidelines algorithm.

In a further extension, to suppress a whole viewpoint, rather than just the voters who voted for an item, we could explore ways to identify those users who share the item's viewpoint and sideline all of them for the selection of the next few items. One way to identify such users would be by clustering: those in the same cluster as those who voted for the just-selected item would be sidelined. Another way would be to use a recommender algorithm: those who are predicted to like the just-selected item would be sidelined.

Despite these limitations and promising directions for future work, it is worth noting one major advantage of the existing Sidelines algorithm: it depends only on the votes for current set of items, and not users' past voting histories or external classifications of the users or items in terms of group affiliations. This characteristic of the Sidelines algorithm is particularly valuable in situations in which more extensive history of user votes or some classification of the voters' or items' viewpoints is not available, such as when users first join a system or when there is a new topic on which previous divisions of people into opinion groups are no longer valid. Simply putting voters on the Sidelines for

a few turns had a noticeable effect in increasing inclusion and the proportional representation of the result sets.

There are other approaches that, like Sidelines, do not require users' voting history or a classification of users. One such approach would be to instead focus on maximizing the representation of additional users. This would be similar in principle to Carbonell and Goldstein's concept of *maximal marginal relevance*, *MMR* (1998). *MMR* reorders result sets after comparing documents' cosine similarities, so that each additional document in the result set is the one that adds the most relevant information not already in the result set, hence the name of the approach. A greedy algorithm could select, for each subsequent item, the item that maximizes the marginal inclusion or minimizes the marginal alienation.

Another direction for algorithm development is to compute what would be the next selected item, with sidelining, simultaneously for all items, using graph traversal techniques inspired by *PageRank* (Brin & Page 1998). For example, Jeh and Widom propose a method, *SimRank*, to compute the pairwise similarities between all pairs of items by traversing a graph consisting of a node for each pair: the similarity score is interpretable as the expected distance that random surfers starting at the two items would travel before meeting each other if they followed random links in the original graph (2002). One idea for our problem would be to construct a bipartite graph with users and items as nodes. Links from users to items would be the votes of the users, with time-weighted scores. Links from items to users would be included, with weights inversely proportional to the similarity as computed with *SimRank*. Then, a fixed point of the graph flow could be computed in order to generate a rank score for each of the items; the items with the highest scores would be chosen for the output set.

In another graph traversal approach, Mei et al proposed the *DivRank* algorithm (2010). They note that traditional graph traversal techniques focus on prestige (or

centrality), generally without consideration of diversity, despite general agreement that breadth of items is often important alongside the prestige or fame of the items in a set. In contrast to *PageRank*'s time-homogenous random walk, *DivRank* uses a vertex-reinforced random walk that creates forward feedback over time (fame begets more fame). Over time, connected vertices compete for prestige, and high-prestige vertices absorb the prestige of their neighbors, thus, in the ultimate ranking, the top-ranked vertices often have lower connectivity – that is, they are farther from each other – than the top-ranked items from *PageRank*. In our problem, *DivRank* could be used with user-votes as edges between different items: an edge is drawn, and weighted, between two items according to the number of users who voted for both items.

It would be valuable to conduct an experiment comparing the impact on our diversity metrics of *MMR*, *DivRank*, and *Sidelines*. Such a study, though, is beyond the scope of this dissertation.

Conclusion

Opinion diversity, although a desirable feature from a societal standpoint and from the standpoint of at least some individual readers, may not naturally occur when the most popular items are selected. The Sidelines algorithm, which suppresses voter influence after a preferred item is selected, is one way to increase diversity. In our experiments, it provided modest increases in diversity according to several different metrics. Perhaps most strikingly, it reduced the tipping toward the majority group even though it operates without any information about the group affinities of people or items. Opportunities remain for research on improved algorithms that take into account additional information, such as group affinities or past voting histories.

CHAPTER III

Sidelines Experiment

To see if readers would notice the difference between the Sidelines and Pure Popularity algorithms described in Chapter II, we conducted a web-based experiment from 27 October to 1 December 2008 using the blog links as votes. In this study, we presented subjects with a list from either the Pure Popularity or the Sidelines algorithm. Because we designed the Sidelines algorithm to increase the diversity of items in the result set, and because our work in Chapter II showed that Sidelines is more resistant to a "tipping" phenomenon, we had several hypotheses related to the diversity goals of the Sidelines algorithm that we wanted to test empirically. In particular, we believed that Sidelines-generated collections would increase subjects' chances of finding challenging and surprising items, and that subjects would have greater variation in their agreement with the items in the collection.

Topic		Question	Findings and outcomes
News aggregators	Selecting diverse collections	• Do people notice the difference in diversity between collections generated by Sidelines and by Pure Popularity?	• Yes, they do.

Table III-1 Summary of questions and findings in this chapter.

Study design

We recruited subjects using a local email list and links posted to the researchers' Facebook profiles. Subjects visiting the site were randomly assigned to see the current snapshot of results from one of two algorithms (they were not told which, or indeed that there were two), asked for reactions to each of the items, and then for reactions to the set as a whole. 40 subjects completed the survey; one response was discarded, as the subject had responded to the questions without actually reading any of the links.

Each subject first was shown a list of links to 12 items, generated by one of the algorithms within the previous 30 minutes. Below each link, we asked subjects to respond to three questions about that item (Figure III-1): did they think the facts were true, did they agree with the opinions presented by the author of the linked item, and had they seen or heard about the story before participating in our survey?

Figure III-1 Portion of the online survey asking users to read links and reply to questions.

We then asked readers to respond to the collection as a whole. On a 5-point scale, did the collection seem liberally or conservatively biased? How complete a range of political opinions did they feel it included (5-point scale from very complete to very

incomplete)? Did they find opinion-affirming or opinion-challenging items? Did they find something surprising?

Finally, we asked subjects about themselves and their preferences. How did they value opinion diversity, topic diversity, opinions that agree with their own, and credibility of facts in news aggregator results? What were their political preferences, on a 7-point scale from extremely liberal to extremely conservative and on a 7-point scale from strong Democrat to strong Republican?

Hypotheses about the *Sidelines* results

As one way of measuring opinion diversity, we calculated the variance in how much each subject agreed with the opinions in each article in their result set. A higher variance would indicate a more diverse result set, so we expected that subjects who viewed a list of links generated by the Sidelines algorithm would have a higher variance in how much they agreed with the opinions presented in each article in the collection than subjects who viewed results from the Pure Popularity algorithm:

H1. Subjects viewing Sidelines algorithm results will report higher variance in their agreement with the opinions in articles presented than subjects viewing the Pure Popularity algorithm.

As our sample of voting blogs was somewhat liberal, and we anticipated a mostly liberal set of respondents (due to our recruitment method, discussed above), we hypothesized that the Sidelines algorithm would result in a greater chance of challenging and surprising items.

H2. Subjects viewing a list of links generated by the Sidelines algorithm would be more likely to find something challenging than subjects viewing results from the Pure Popularity algorithm.

H3. Subjects viewing a list of links generated by the Sidelines algorithm would be more likely to find something surprising than subjects viewing results from the Pure Popularity algorithm.

Because of the Sidelines algorithm's resistance to tipping, we believed respondents would feel that the Sidelines results represented a more complete and less biased range of items than results from the Pure Popularity algorithm.

H4. Subjects viewing a list of links generated by the Sidelines algorithm would rate the collection of links as including a more complete range of political opinions than subjects in the Pure Popularity group.

H5. Subjects viewing a list of links generated by the Sidelines algorithm would rate the bias of the collection as more neutral than subjects in the Pure Popularity group.

Given Stromer-Galley's findings that people say they seek out diversity (2003), we also expected the Sidelines algorithm would lead to result sets that people at least said were more satisfying than the Pure Popularity results, after we had primed them to think about the diversity of the result sets.

H6. Subjects viewing a list of links generated by the Sidelines algorithm would report being more satisfied with the collection of links than subjects in the control group.

The substantial counter-evidence, in favor selective exposure preferences in online news readership, would also support an alternative hypothesis that people would be less satisfied with the Sidelines collections.

Results

First, we report on which attributes of news aggregator collections subjects reported valuing. Subjects reported that they value diversity of opinions more than agreeable opinions (paired Wilcoxon signed-rank test, $V=49$, $p < 0.05$), consistent with Stromer-Galley's findings (2003). Subjects indicated valuing credibility of facts more than either topic diversity (paired Wilcoxon signed-rank test, $V=265$, $p < 0.001$) or opinion diversity

(paired Wilcoxon signed rank test, $V=361$, $p < 0.001$), and valuing topic diversity more than opinion diversity (paired Wilcoxon signed rank test, $V=161$, $p < 0.05$).

	Both	Pure Popularity	Sidelines	p-value
Value diversity in opinion (1 not at all, 5 very much)	3.64	3.30	4.00	0.065
Value diversity in topic (1 not at all, 5 very much)	4.05	4.05	4.06	0.729
Value opinion agreement (1 not at all, 5 very much)	2.77	2.75	2.79	0.844
Value credibility in facts (1 not at all, 5 very much)	4.64	4.45	4.84	0.237
Liberal to conservative (1 extremely liberal, 7 extremely conservative)	2.30	2.56	2.06	0.485
Party affiliation (1 strong liberal, 7 strong conservative)	2.22	2.55	1.92	0.179

Table III-2 Subject groups for Pure Popularity and Sidelines experiment. Wilcoxon rank sum test.

Table III-3 summarizes the results from our online experiment. Overall, the only significant (either statistically or clinically) difference was in the likelihood of finding a challenging item in the collection: viewers of Sidelines collections were 1.78 times as likely to find a challenging item as the viewers of Pure Popularity algorithm. I next discuss the results as they relate to each hypothesis.

	Pure Popularity	Sidelines	Test used	p-value
Variance in opinion agreement	1.08	1.38	t-test	0.233
Variance in credibility	0.65	0.83	t-test	0.210
Credibility (1 not credible, 5 credible)	3.65	3.71	Wilcoxon rank sum	0.844
Overall bias (1 conservative, 5 liberal)	3.40	3.16	Wilcoxon rank sum	0.486
Completeness (1 very incomplete, 5 very complete)	2.50	2.95	Wilcoxon rank sum	0.153
Satisfaction (1 very unsatisfied, 5 very satisfied)	2.60	2.63	Wilcoxon rank sum	0.940
Found something affirming (0 no, 1 yes)	0.94	0.95	Binomial test	1.000
Found something challenging (0 no, 1 yes)	0.50	0.89	Binomial test	***<0.001
Found something surprising (0 no, 1 yes)	0.68	0.68	Binomial test	1.000

Table III-3 Results of online experiment based on political blog links.

H1 Opinion Diversity. Though the values are in the expected direction – higher variance in agreement for the Sidelines – the result is not statistically significant, and the effect size would be fairly small as well (Table III-3, row 1). Thus, H1 is not supported.

H2 Challenge. The subjects, who mostly self-identified as Democrats and liberals, were more likely to find something that challenged their opinions in the Sidelines result sets; an 89% chance of finding something challenging vs. 50% (Binomial test, $p < 0.001$; Table III-3, row 8).

H3 Surprise. H3 is not supported; there was no difference in the likelihood of a subject finding a surprising item in the Sidelines or Pure Popularity result sets (Table III-3, row 9).

H4 Completeness and *H5 Bias.* The subjects reported that that the Sidelines modification may have delivered slightly more neutral (Table III-3, row 4) and slightly more complete (Table III-3, row 5) results sets with regard to range of political opinion.

While these indicators are directional, any effects, if present, are rather small and not statistically significant.

H6 Satisfaction. There is no apparent difference in subject satisfaction between the two algorithms (Table III-3, row 6). This is an important result, as the subjects viewing Sidelines collections were not less satisfied than subjects viewing the Pure Popularity collections – despite being much more likely to have their opinions challenged by an item in the collection.

Discussion

A free-response section of the experiment's satisfaction question indicated that other shortcomings of our aggregator – particularly topic redundancy – drove satisfaction down in both conditions. Unlike many popular news aggregators (e.g., *Google News, Memeorandum*), we did not cluster similar articles. On days when one particular news story was receiving substantial attention, notably right around the Presidential election, when most of our subjects completed the survey, the list of 12 items presented to survey respondents might contain many links to different coverage of the story; many subjects complained about this in the free response section of the survey. Without the inclusion of these news aggregator features, it is difficult to assess how the Sidelines algorithm affects satisfaction with result sets. While it is encouraging that the more diverse collections produced by Sidelines did not decrease satisfaction, it is possible that any effects of diversity on satisfaction were crowded out by other factors affecting satisfaction. Future iterations of our aggregation algorithms – both the baseline we use for comparison and those designed to promote diversity – should cluster related stories by topic to reduce redundancy in the result sets.

While Sidelines was designed for news aggregator sites that rely on user votes, there are likely more, and perhaps better, applications. As seen in this experiment and in the evaluation presented in Chapter II, it can also produce more diverse and

representative result sets from collections of links, by treating each link as a vote. Other voting sites might also benefit from the Sidelines algorithm. *Google Moderate*, for example, allows people to submit questions or ideas and for the site's other users to vote on these ideas and reply to them. When the Obama transition team used an instance of this application in 2008, the front page became dominated by questions about the legalization of marijuana, since interested users could vote up all questions related to this idea. If *Google Moderate* had used the Sidelines algorithm, would it have surfaced a more representative set of issues? Could Sidelines also provide benefits for companies, such as Threadless, that decide what products to produce based on user votes and ratings?

Conclusion

In this experiment, using real-world link data and news items, the Sidelines algorithm produced result sets in which subjects were nearly 1.8 times as likely to report finding an opinion-challenging item as in the Pure Popularity collections. Thus, not only does Sidelines produce more diverse result sets according to the metrics discussed in Chapter II, this increase in diversity is also subjectively noticeable to viewers of the collections. Furthermore, the algorithm achieved this increase in diversity without a decrease in satisfaction with the presented collections of items.

CHAPTER IV

Presenting Diverse Political Opinions: How and How Much[4]

If designers of news aggregators turn to the research literature to learn how much opinion diversity people seek or tolerate in their political news, they will be confronted with a range of competing theories and evidence. According to different theories, readers may seek out diversity, they may avoid it, or they may seek reinforcement and tolerate challenge only when accompanied by sufficient reinforcing information.[5] Each of these alternatives has different design implications.

As discussed in the introduction, there are competing theories and evidence about people's preferences and behavior for viewpoint diversity in online settings. Selective exposure theory predicts that people both seek out affirming items and avoid challenging, or counter-attitudinal, items (Frey 1986, Mutz & Martin 2001). We will refer to this as the *challenge-aversion* hypothesis. People who are challenge-averse would prefer news and opinion aggregators that display only agreeable content. If people are challenge-averse, it will be difficult for designers to create single collections that appeal to audiences with

---

[4] This chapter is an expanded version of Munson, SA; Resnick, P. 2010. "Presenting Diverse Political Opinions: How and How Much," *Proceedings of CHI* 2010.

[5] In this chapter, I focus on preferences for exposure to diversity in online news reading. Esterling et al (n.d.) conducted a similar study to explore individuals' preferences for exposure to diversity in deliberative environments. They review and work to distinguish between four typologies: *disagreement tolerant* (in which satisfaction diminishes slightly as disagreement increases), *disagreement phobic* (in which satisfaction decreases sharply as disagreement increases), *disagreement philic* (in which satisfaction increases disagreement seeking), and *curious* (in which satisfaction increases, up to a point, and then decreases as disagreement increases).

diverse opinions. Moreover, in personalized presentations, people will prefer homogeneous collections of all agreeable items. In either case, it will be hard to meet the public policy goal of high exposure to challenging information, unless presentation techniques can be developed that make that exposure more palatable.

Other arguments and studies dispute the challenge aversion theory and offer a contradictory hypothesis of *diversity-seeking* (Stromer-Galley 2003, Horrigan et al 2004, Kelly et al 2005). The preferences of diversity-seeking individuals are consistent with public policy goals; they would be most satisfied with collections that contain a range of views. If people are diversity-seeking, it may be possible to simultaneously satisfy an audience with diverse viewpoints. Moreover, diversity-seeking people would also choose personalized collections with much less than 100% agreeable content.

A third hypothesis is that people are *support-seeking* (Garrett 2009). Like challenge-aversion, this hypothesis posits that people seek out affirming items but rejects the idea that they avoid challenging items. Support-seeking individuals would prefer news aggregators that show them a sufficient amount of agreeable content, and would be indifferent as to whether items beyond that amount support or challenge their views. Thus, with a large number of items in a collection, it would be possible to please people with a variety of viewpoints, and in an individualized collection, people would not mind the inclusion of additional challenging items.
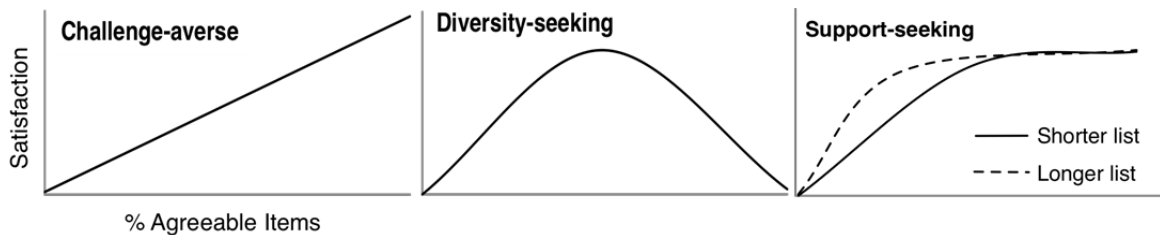
Figure IV-1 Competing hypotheses about preferences for agreeable and challenging items.

In this study, we presented readers with a list of political opinion stories, with varying numbers of challenging and agreeable items, and then measured their satisfaction with the list, to distinguish among these different theories about people's preferences for

agreeable and challenging information. We would expect to see different relationships between the percentage of the list that is agreeable or disagreeable and reader satisfaction, as summarized in Figure IV-1. If people are challenge-averse, then we would expect that a higher percentage of agreeable items and lower percentage of disagreeable items would be preferred. On the other hand, if people prefer diversity, then we would expect them to be most satisfied when the list contains both agreeable and disagreeable views. Finally, if people prefer agreeable items but are not particularly averse to challenging items, then the count of agreeable items, not the percent of items that are agreeable, should drive satisfaction. Thus, satisfaction should increase toward the asymptote as the percentage of agreeable items increases, and should approach the asymptote at a lower percentage for longer lists.

If there are people who are challenge-averse or support-seeking, this opens a new question: can we use presentation techniques to make people who prefer support equally satisfied with a lower percent or count of agreeable items? To assess this, we tried highlighting agreeable items, as well as placing them first in the collection.

Topic		Question	Findings and outcomes
News aggregators	Diversity preferences	• With what percent of agreeable and disagreeable items are people most happy?	• Some people seek diversity, while others avoid challenge.
	Presenting diverse collections	• Does highlighting agreeable items increase satisfaction with collections that include challenging items? • Does highlighting agreeable items and placing them first increase satisfaction with collections that include challenging items?	• Not overall. Highlighting agreeable items made peoples' reactions stronger. • No. Highlighting agreeable items and placing them first made people less satisfied overall.

Table IV-1 Summary of questions and findings in this chapter.

Exposure to diversity

Prior research on exposing individuals to diversity in the news has discussed both the selection and presentation components of the problem. Park et al developed and evaluated the *NewsCube* system (2009). *NewsCube* classifies content into different viewpoints or "aspects" on issues. Subjects who participated in a *NewsCube* trial read more articles about a controversial issue than subjects who used *Google News* to learn about the same topic. The articles read by *NewsCube* subjects appeared to cover a greater breadth of opinions (per reader). Subjects who said they would not normally compare different articles on a topic said that the aspect-presentation made them want to read more articles. This suggests that certain presentations of information can increase the diversity of opinions that participants are motivated to access, even if some of those opinions are disagreeable.

Other work also examines both the presentation and selection questions. Oh et al built a blog search engine that classified results according to political viewpoints (liberal or conservative) and then used that information in the presentation of search results (2009). Results were either labeled or sorted into two labeled columns according to this bias. Subjects had mixed reactions to these presentations, but generally preferred seeing a column of liberal items and a column of conservative items. Those dissenting preferred to decide for themselves what was liberal or conservative, or felt that the labels added too much polarization to the results – even though they did not change which results were shown. The researchers also found that in the two-column layout, liberal sources accounted for a greater portion of the liberal searchers' clicks (there were too few conservative subjects to observe trends among these subjects).

In Chapter II, I proposed an algorithm – based on user votes – for selecting diverse items and discussed three metrics for evaluating selection algorithms: inclusion, alienation, and proportional representation. While these metrics provide insight into algorithms' performance at including items for which many users voted, and how well the

proportions of items align with proportions of users, they are incomplete as measures of the diversity goals discussed above. That is, while they tell us about whether the collection includes items that represent voters, these metrics do not tell us whether the voters *feel* represented. Many of the desirable and undesirable outcomes depend not on whether diverse results are present but instead on viewers' *reactions* to that diversity. For example, even if ideas are represented in a collection in the same proportion with which they are held, people holding minority opinions may not notice the small number of items representing their views among the many more that challenge their opinions. This may cause them to seek out sources where they can more easily find items that affirm their views, leading to polarization. It is important, then, to understand reactions to different levels of supporting and challenging items, and to identify design choices that may affect viewers' reactions to these collections.

Methods

All experiments followed the same general design: we provided each subject each day with a list of 8 or 16 links to opinion articles, with known biases, about United States politics. The articles were found on blogs or in the mainstream media. Each item was displayed as an article title together with its first or last paragraph, as shown in Figure IV-2. The selection and presentation of items varied among subjects.

Figure IV-2 Example articles and question form.

Subject recruitment

We recruited subjects from Amazon.com's Mechanical Turk service, a system that allows remote workers to complete small tasks for small payments. Mechanical Turk has been used by other researchers for annotating data, and Kittur et al have published guidance for using Mechanical Turk in research (2009). These guidelines were useful in planning our study.

We used a Qualification test, for which subjects were not paid, to initially screen the Mechanical Turk workers. We asked subjects about their location, age, political knowledge, and political preferences. We only accepted Mechanical Turk workers who self-reported (to Amazon for payment purposes) a United States location and whose

previous task approval rate exceeded 90%. Additionally, we screened each Mechanical Turk worker for some basic US political knowledge using multiple-choice questions:

1. Who is the current Vice President?

2. Which party is George W. Bush a member of?

3. For what position is Sonia Sotomayor currently a nominee? (And later, to what position has Sonia Sotomayor been appointed?)

Potential subjects had to correctly answer two of the three questions to participate, though on average, accepted subjects answered 2.98 questions correctly.

We also asked potential subjects about their party affiliation (7-point scale from strong Democrat to strong Republican) and about their political preferences (7 point scale from strong liberal to strong conservative). We selected for subjects whose party affiliation matched their liberal or conservative preferences (for example, we screened out subjects who reported being both a strong Republican and a liberal). Furthermore, because we wanted subjects for whom we could predict an article would be agreeable or challenging with some confidence, we also filtered out subjects who were more neutral or independent. To be included, subjects had to have a mean position (as the average of their responses to the two seven-point scale questions) of ≤3 or ≥5.

Although subjects were not a random sample of the U.S. population, the subjects were diverse in geography, age, and gender. Subjects lived in 37 of the 50 U.S. states (Figure 3). Their mean age was 34.3 years (median: 31 years, standard deviation: 11.8 years). 83 were men, 87 women.

Figure IV-3 Locations of Mechanical Turk workers who participated in the study.

Item selection

We selected items based on links from a panel of 500 political blogs – the same set as presented in Chapter II. Each day, we selected the 40 most popular liberal and conservative items, based on number of links to the items from liberal and conservative blogs in the previous 36 hours. Items were defined as conservative if the ratio of the probability of any conservative blog in our panel linking to the item compared with the probability of any liberal blog in our panel linking to the item was at least 2:1, and vice-versa for liberal items. The selection system also filtered out tweets, Twitter accounts, Wikipedia articles, and YouTube videos. Before including the selected articles in our pool of items, each morning researchers manually inspected each candidate item and removed items that did not match the predicted bias (e.g. a liberal item coded as "conservative" because conservative bloggers linked to it to highlight a disagreement with liberals). We also removed items that did not contain or report on opinion, as well as posts that contained only video, images, or audio. On average, this left 23 articles of each bias, per day.

30 turkers were assigned to a manipulation check survey. In this survey, each turker was presented with a list of three links and asked to what extent they agreed with each link on a 5-point scale. This was run early each day, after the researchers reviewed the list of items. Based on raters' responses, we then removed items that were not found to

match their predicted bias. An item that was predicted to be liberally biased would be removed if liberal raters did not agree with it, or if conservative raters did agree with it. Raters agreed with our predicted ratings (as supporting or challenging the raters' opinions) 74% of the time. These disagreements were not distributed evenly across items, and 16% of the items were removed because of disagreement.

We did not require all items to get 100% agreement for inclusion, as we do not expect parties to be 100% united in opinion. Initially, an item was queued for 3 ratings (including at least 1 from a participant from each party). If no raters diverged from the prediction, an item was included. If two or more raters diverged from the prediction, we discarded the item. If one rater diverged from the prediction, the item was queued for an additional rating from someone with the same party affiliation. If the fourth rater disagreed with the prediction, the item was discarded; if this rater agreed, the item was included. In cases when this rating was not obtained (insufficient participation in our Mechanical Turk task), the item was discarded. On five days, this removal resulted in too few liberal or conservative items, in which case the system filled in with items from the previous day as needed.

Experimental design

The subjects not assigned to the manipulation check condition viewed a list of items and were asked one of two questions about the representativeness of the collection as a whole. Subjects were randomly assigned to one of six experimental conditions in a 2x3 factorial design:

1. *Total number of items:* 8 or 16

2. *Presentation:* a list with agreeable and disagreeable items interwoven, a list with agreeable and disagreeable items interwoven and the agreeable items highlighted, or a list with the agreeable items first and highlighted followed by the disagreeable items.

Repeated measures were collected: each subject could complete one survey per day, based on the items selected that day. Subjects remained in the same experimental condition throughout the study, but the number of agreeable items was randomly chosen for each subject each day.

The instructions read:

The following list contains some of the most-linked to political opinion stories from the last few days. Please look at the list as you might if you were to visit a website like Digg or Reddit (you may click on and read as many or few as you like). Then answer the questions at the bottom of the page. Thank you!

Additionally, subjects in the agreeable first or highlight conditions were told that items they were predicted to agree with would appear highlighted in the list.

Subjects were also assigned to one of two questions. The first question (44 subjects) was our primary outcome measure and asked about the subject's satisfaction with the range of views:

"Suppose this was the front page of a political opinion aggregator. How would you feel about the viewpoints represented in it?" (5 point Likert-like scale, Very dissatisfied to very satisfied)

The second question (39 subjects) asked about the bias of the collection:

"What, if any, is the political bias of this collection?" (5 point Likert-like scale, Very Liberal to Very Conservative)

We used this question to help us understand our findings. If a particular presentation feature affected satisfaction, is it because it changed how the subjects perceived the collection's bias, or is it because the subjects simply did not like the feature?

In addition to these questions, each time a subject viewed a list, they were randomly asked either to provide a free-text explanation for why they gave the rating they did or to repeat a question from the pre-test (party affiliation, liberal to conservative, age, or gender). The free response question helped us to understand why subjects gave the ratings they did and if they were interpreting the questions as intended. Repeating

questions from the qualification test follows a recommendation from Kittur et al to ask verifiable questions (2009). Five subjects (four from the satisfaction question, one from the bias question) changed their answer substantially (e.g. aging more than one year or in reverse, changing gender, or shifting on either of the political spectrum questions by two points or more). Though there are many possible explanations for these shifts – such as shared accounts within a household, careless clicking, easily shifting political opinions, deliberate deception, or lack of effort – all of these explanations are not desirable for study subjects, and so these subjects and their responses were excluded from our analysis, leaving us with 108 subjects (40 responding to the satisfaction question, 38 responding to the bias question, and 30 rating articles for the manipulation check).

After a five-day warm-up period with variable pay (to identify an appropriate price), subjects were paid $0.75 for rating a collection of items. This pay may seem high compared to some expectations for Mechanical Turk labor. We believe that we had to pay a higher price because our task both required successful completion of a qualification and was only available once per day. Interestingly, many more people (171) completed the qualification test and were approved than actually returned to complete a task.

We collected data daily from 22 July – 14 August, and then on alternating days from 26 August to 10 September, 2009. During this time period, the major topics in articles displayed included national debates about healthcare reform and the Waxman-Markey cap and trade bill, the death of Massachusetts Senator Edward Kennedy, discussion about the success of the Troubled Asset Relief Program and the American Recovery and Reinvestment Act (including the "Cash for Clunkers" program), the release of two American journalists from North Korea, continued discussion about political unrest in Honduras, the abrupt resignation of Alaska Governor and former Vice Presidential candidate Sarah Palin', and speculation about 2010 Congressional and gubernatorial elections. Previous selective exposure studies have been criticized for being short term laboratory studies that might be particularly subject to social desirability bias

and other confounds (Taber & Lodge 2006, Knobloch-Westerwick & Meng 2009), and so we allowed subjects to return up to once per day, on all days during which we collected data, with the hope that a more longitudinal design, featuring a range of current political issues on different days, would more accurately reveal subjects' preferences than a one-off laboratory study.

A previous study found that Mechanical Turkers' efforts were not tied to the amount of payment (Mason & Watts 2009). On average, our subjects rated a list in 6.3 minutes (5.8 minutes for 8-item lists; 7.3 minutes for 16-item lists). No subjects completed the task in less than a minute. For comparison, *Alexa* (accessed 16 September 2009) reports that *Digg* visitors spend an average of 4.2 minutes per day on the site, *Reddit* visitors spend an average of 6.3 minutes, and *Memeorandum* visitors spend an average of 2.0 minutes.

Results

I next discuss the results of our study with respect to individuals' diversity preferences and to our efforts to nudge challenge-averse individuals to tolerate more opinion diversity using simple presentation techniques.

Diversity preferences

In our first look at the data, we found that when the list contained a low percentage of agreeable items, almost all subjects were very dissatisfied. When a high percentage of items were agreeable, however, there was greater variance in responses: some subjects were very satisfied, some subjects were very dissatisfied. This suggested that there may be individual differences, with some people diversity-seeking and some challenge-averse.

To confirm this, we analyzed the open-ended responses for the question about why subjects in the satisfaction question group gave the ratings they did. Some subjects wrote that they specifically did not want a list of solely supportive items and that they

want opinion aggregators to represent a fuller spectrum of items, even if that includes challenging views. Two researchers coded all free-text responses for similar remarks and then coded participants as diversity-seeking if they had made at least one such comment. Our standards were strict: subjects had to write that they either

1. wanted a fuller spectrum of views even though their views were represented in the majority of items in the list, e.g.,

    "It all seems liberal. I'm liberal, but I think it's good to get dissenting opinions instead of having all the articles slanted the same way. I'd really like seeing pro and con articles on some of the topics."

    "The articles in this list showed some of both sides on some issues, but on other issues like health care was rather one sided. If that and a few other articles had been given two sides I would be completely satisfied. I like to read both sides even though I am mostly conservative."

    or

2. were pleased with the balance of items and would not want more supporting items, e.g.,

    "There is an even distribution of right and left wing articles. I think it is best to cover both sides of the issue."

    "I like that there are views from both Democrats and Republicans and seems to be a great mix of both sides of the fence."

To avoid potential biasing of the coders, coders were not informed of the actual number of agreeable items presented to a subject or of the subject's satisfaction score for the items when coding the subject's explanatory comments. We did not use the actual properties of the list associated with a comment when we coded, out of concern that we would code people as diversity seeking when the subject's remarks were ambiguous in order to explain behavior. Our inter-rater reliability, calculated as Cohen's kappa (1960), was 0.89. All raters coded all subjects. Landis and Koch (1977) characterize agreement

above 0.8 as "almost perfect agreement." We decided the disagreements through discussion.

Ten out of the 40 subjects in the satisfaction condition were coded as diversity-seeking (25%). This is likely an undercount given our strict coding criteria and that some participants never saw a collection that would prompt different reactions from diversity seeking participants. Once we separated out the diversity-seeking individuals, our results were much clearer (Figure IV-4). The linear regression predicting satisfaction based on percent agreeable items, whether a subject is diversity seeking, and the interaction terms between these (to the quadratic polynomial) shows significant interaction effects between percent agreement and diversity seeking for predicting satisfaction (Table IV-2). Because subjects were able to rate multiple collections (one per day), in all regression results we cluster responses by subject, which reduces degrees of freedom and inflates standard error estimates, to correct for correlation of repeated measures.[6]
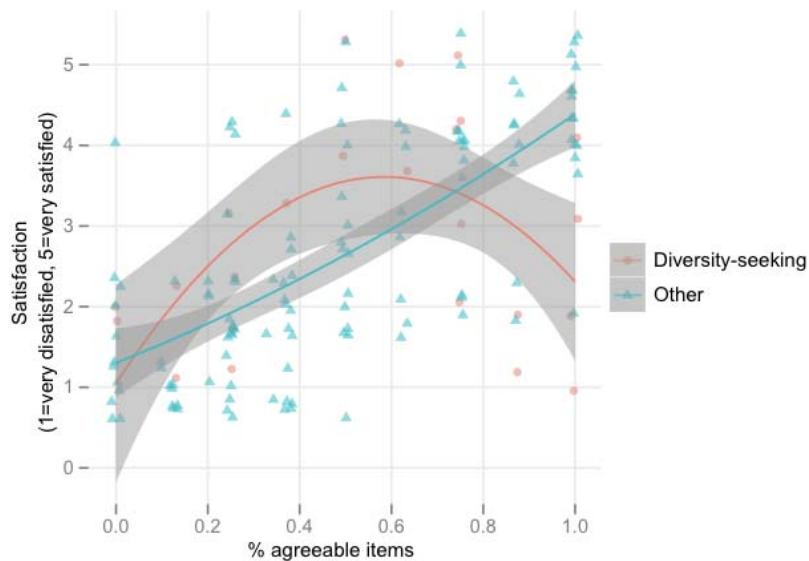
Figure IV-4. Comparison of satisfaction at different percentages of agreeable items for diversity-seeking and other (either challenge-averse or support seeking) individuals. Fit lines according to regression model in **Table IV-2**. The grey band includes ± one standard error of the prediction.

---

[6] All regressions were performed using STATA 10's `regress` command with the robust cluster by subject option.

	β	Std Err	p-value
Intercept	1.30	0.23	<0.001
% Agreement	2.28	0.76	<0.010
(% Agreement)^2	0.80	0.66	ns
Diversity seeking	-0.25	0.63	ns
% Agreement * Diversity seeking	6.49	3.16	<0.050
(% Agreement)^2 * Diversity seeking	-8.32	3.11	<0.050

Table IV-2 Linear regression results for satisfaction (1-5). n=145 from 40 subjects, $F(5,39)$ = 29.63 (p <0.001); adjusted R^2 0.4776.

SUPPORT-SEEKING OR CHALLENGE-AVERSE?

We then examined whether the remaining 30 individuals were support-seeking or challenge-averse. Challenge-averse subjects would be equally satisfied at the same percentage of agreeable items, regardless of the length of the list. Support-seeking subjects, in contrast, would be equally satisfied at the same number of agreeable items, regardless of the length of the list. If we were to find evidence that people are support-seeking, it would offer one simple strategy for addressing the public policy challenge of exposing people to more perspectives. Simply presenting a longer list of results could lead to sufficient counter-attitudinal information exposure, so long as people did not become fatigued with a longer list.

In our analysis, we did not find evidence of support-seeking individuals. Table 2 presents the linear regression model for:

satisfaction = β_1 + β_2(% agreeable items) + β_3(listlength$_{16}$)+ β_4(listlength$_{16}$ * % agreeable items)

In this model, listlength$_{16}$ is a dummy variable equal to 1 for 16-item lists (longer lists) and 0 for 8-item lists (shorter lists).

	β	Std Err	p-value
Intercept	1.61	0.33	<0.001
% Agreement	2.69	0.40	<0.001
List length = 16	-0.58	0.40	ns
% Agreement * (List length = 16)	0.53	0.53	ns

Table IV-3 Linear regression results for satisfaction (1-5) with data from diversity-seeking subjects withheld. n = 112 from 30 subjects, $F(3,29)$ = 50.43 (p <0.001); adjusted R^2 0.5079.

We exclude data from the 10 diversity-seeking individuals. If there are any effects of list length, they are in the opposite direction of what would be expected for support-seeking behavior. In a plot of the percent agreeable items and satisfaction (Figure IV-5, top), the slope of the fit lines for the two list lengths follow each other quite closely, suggesting that count does not matter. When we plot the number of agreeable items (Figure IV-5, bottom), we can see a clear divergence. Furthermore, two agreeable items out of a total of eight is superior to two agreeable items out of a total of sixteen ($t(7.373)$ = 3.3471, $p<0.05$). It is possible that subjects became fatigued with longer lists, particularly longer lists of challenging items, confounding our results, though the list was not longer than that found on many news aggregator sites – the *Digg* front page shows 18 items and the *Reddit* front page shows 25 items – so we do not believe that length drove satisfaction down. The presence of challenging items, not just the count of agreeable items, shapes satisfaction. We conclude that the remaining subjects as a group are challenge-averse, though it remains possible that a few individuals are support-seeking.
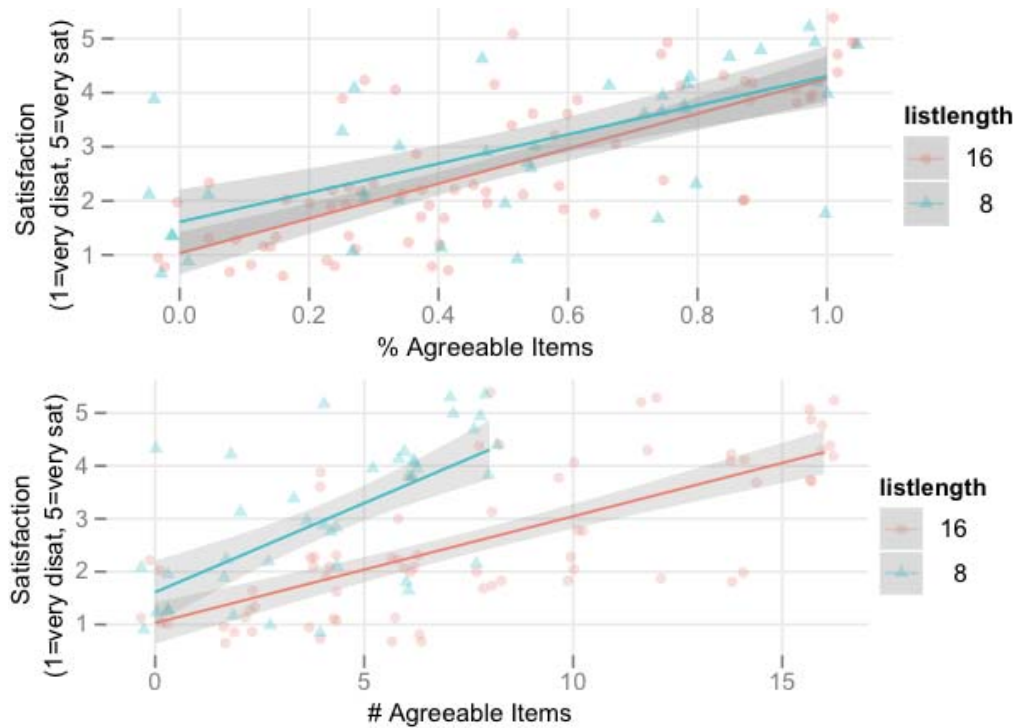
Figure IV-5 *Top*: **Percent agreement and satisfaction for 8 and 16 item lists.** Fit lines from model in Table 2. *Bottom*: **Number of agreeable items and satisfaction for 8 and 16 item lists.** Fit lines: satisfaction = β_1 + β_2(# agreeable items) + β_3(list length)+ β_4(list length * # agreeable items). Responses from individuals coded as diversity-seeking excluded from both plots.

We also looked to see if subjects' demographic characteristics were correlated with whether they were diversity-seeking or challenge averse. With this small number of subjects, we saw no statistically significant differences in age, gender, partisan or political preference, or partisan or political extremity between the two groups. We do, note, however, that the challenge-averse subjects were 57% female while the diversity seeking subjects were 80% female; future studies should explore whether this proportion difference actually exists in the larger population or occurred by chance in our study.

Presentation techniques for challenge-averse readers

Having found that at least some people exhibit challenge-averse behavior, we investigated whether presentation techniques could increase their satisfaction with collections that contained some challenging items. Table IV-4 presents a linear regression model for the

effects of presentation style and their interaction effects with percent agreement. We will discuss the effects of each presentation style separately.

	ß	Std Err	p-value
Intercept	1.59	0.29	<0.001
% Agreement	2.60	0.36	<0.001
Highlighting only	-0.60	0.41	ns
% Agreement * Highlighting only	1.29	0.60	<0.05
Agreeable first	-0.97	0.31	<0.010
% Agreement * Agreeable first	0.64	0.44	ns

Table IV-4. Regression model for a reader's satisfaction (1-5) as predicted by percent agreement and presentation style. (baseline presentation interweaves agreeable items and does not include highlighting). $n = 121$ from 30 subjects, $F(5,29) = 67.42$, $p < 0.001$, adjusted $R^2 = 0.564$.

HIGHLIGHTING ONLY

We had expected that, for challenge-averse individuals, highlighting agreeable items would increase their satisfaction at all percentages of agreeable items, by helping them identify these items even when they were rare in the collection. Figure IV-6 presents results from a reduced model excluding participants in the agreeable first condition.
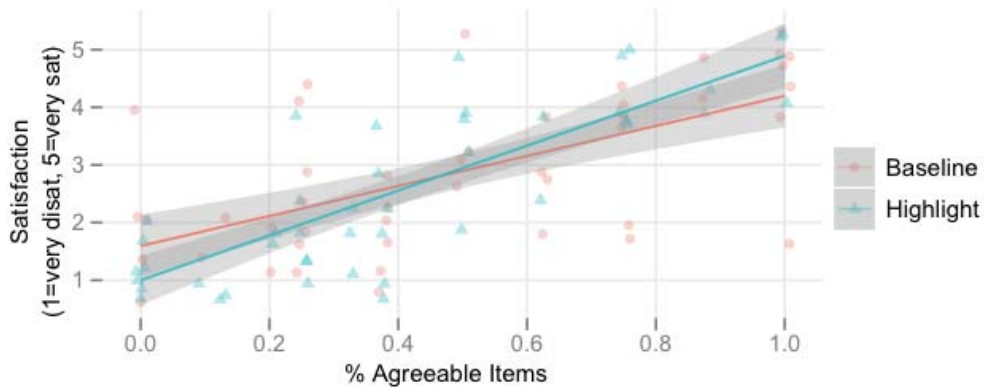
Figure IV-6. Reduced model for highlight -only: satisfaction $= \beta_1 + \beta_2$(% agreeable items) $+ \beta_3$(highlight) $+ \beta_4$(highlight*% agreeable items)

Contrary to our expectations, there is no main effect of highlighting. Instead, there is a significant interaction terms between highlighting and percentage of agreeable items. Highlighting agreeable items makes a reader's reaction – whether it is satisfaction with a

high percentage of agreeable items or dissatisfaction with a low percentage – more extreme.

At the lower range of agreeable items, where we had expected highlighting to have the greatest improvement in satisfaction, subjects' satisfaction is actually decreased. In the mid-range, particularly important for public policy goals of showing viewpoints in proportion to how they are held, highlighting has essentially no effect. With high percentages of agreeable items, it may be possible to highlight agreeable items and include a few challenging items while achieving the same satisfaction as a list of only agreeable items.

Among subjects to whom we posed the question about the collection's bias, we saw a similar interaction effect between highlighting and the percent of agreeable items, suggesting that the effects of highlighting on a subject's satisfaction with the collection is moderated by how highlighting affects their perception of bias in the collection. In other words, it seems that highlighting helps subjects judge the percentage of agreeable items, and the perceived percentage drives satisfaction. Consistent with that interpretation, we note that challenge-averse readers in the highlight condition spent an average of 5.1 minutes per collection, compared to 7.5 minutes for challenge-averse readers who read lists without highlighting. Finally, it is also possible that highlighting the agreeable items had a priming effect that affected subjects' responses to the satisfaction question, without actually affecting their actual satisfaction with the collection.

Highlighting+Ordering: Agreeable items first

We also anticipated that placing agreeable items first would increase challenge-averse readers' satisfaction. Instead, readers who viewed agreeable items as highlighted and at the beginning of the collection reported lower satisfaction (Table IV-4). Figure IV-7 displays a model comparing highlight+ordering and the baseline presentation.
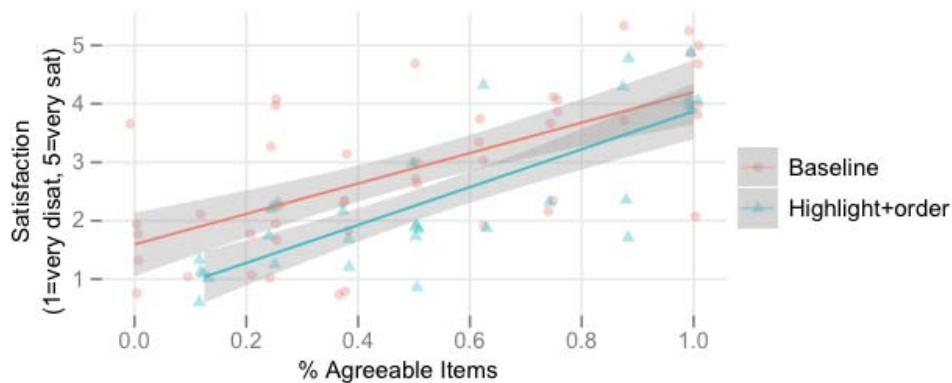
Figure IV-7. Model comparing baseline and highlight+ordering: satisfaction = $\beta_1 + \beta_2$(% agreeable items) + β_3(agreeable first) + β_4(% agreeable items * agreeable first)

The subjects who we asked to rate the collection's bias, however, reported that the collection was more biased in their favor when the agreeable items were highlighted and shown first than when the agreeable items were not highlighted and were interwoven with disagreeable items. This appears to be contradictory, or at least suggests that something other than perceived bias is driving satisfaction in this case. Challenge-averse readers in the highlight+ordering condition spent an average of 5.3 minutes per collection, compared to 7.5 minutes for challenge-averse readers who read lists without ordering and highlighting.

Discussion

Our study finds good and bad news for those with a policy goal of encouraging exposure to a diversity of opinion. The good news is that some people actually prefer collections of items with diverse opinions. They appear not to be the majority, and so it may be important to consciously design specifically for this audience, as they may not naturally be served if designers build applications primarily for the majority, challenge-averse individuals. We do not expect, however, that the proportions of challenge-aversion and diversity seeking individuals we observed generalize to a broader population: among other limitations, our small sample was limited to people with fairly strong political

preferences. A more representative study is needed to better estimate the prevalence of challenge-averse and diversity seeking individuals.

The bad news is that for challenge-averse individuals, designers cannot substitute ordering or highlighting of agreeable items for including more agreeable content. With highlighting, it might be possible to include one or two challenging items in a list of otherwise agreeable items and achieve the same satisfaction from challenge-averse people as with an unhighlighted list that happens to contain completely agreeable items. From the perspective of website operators trying to attract and retain users, this is unlikely to be a desirable tradeoff. It is unlikely to be sufficient challenge to satisfy diversity-seeking individuals, and would leave the news aggregators vulnerable to losing challenge-averse individuals to competitors who offer 100% agreeable items all the time (and hence need no highlighting).

We also cannot rule out that the observed effect of placing agreeable items first is a result of flawed experimental design. By asking about subjects' satisfaction at the end, and placing the agreeable items at the beginning, we may have prompted a recency effect (Broadbent & Broadbent 1981) – that is, their answers were more influenced by the disagreeable items nearer to the question. In an actual political opinion aggregator, truly challenge-averse readers may never scroll that far, while the Mechanical Turk readers may have felt an obligation to read every item because they were being paid rather than reading for their own enjoyment, or they may have skipped looking at the first few items, immediately scrolling down to the questions and looking only at the items closest to the questions.

The responses from the subjects who we asked about the collection's bias, however, appear to contradict the explanation that items nearer the bottom of the list, and thus the question, weighed more heavily in the subjects' consideration of bias and led to the observed decrease in satisfaction with presenting agreeable items first. It is possible that the subjects simply liked the agreeable first presentation less, or, that by revealing we were

willing to change the order of items, we raised expectations and they were frustrated that we did not create a list of entirely agreeable items. Alternatively, subjects may have perceived a fairness norm, which we appeared to violate by changing the order of items in the list, thus decreasing their satisfaction. It is also possible that subjects read the collection differently when we asked them to characterize its bias than when we asked them about their satisfaction with the opinions presented, and that this caused them to experience the collection's bias differently between the two subject groups.

Alternative experimental designs – such as placing the question at the top of the list or having it scroll alongside – may lead to an improved understanding of the effect of ordering. It is also possible that subjects in a lab experiment or Mechanical Turk task will always feel obligated to read an entire list even when they would not do so on an actual website. As with any laboratory experiment, the ecological validity of our study was limited, though its design as an online, longitudinal study hopefully resulted in higher external validity than short-term in-lab studies. Emails from subjects, saying that they had come to value the study's collection of articles as a daily news source, and hoping that it would remain available after the study, support this hope. Nevertheless, field trials examining actual behaviors, such as visits to a site, articles read, and engagement such as comments, are likely to yield measures of preferences and behavior that have higher external validity than a self-reported measure of satisfaction.

We only evaluated a small range of the presentation techniques for lists of items. Alternative ideas that could be evaluated include reducing the space that challenging items occupy in a list by reducing the font size or collapsing challenging items' abstracts. Perhaps such techniques would make palatable a smaller percentage of agreeable items. They would increase the risk, however, that people would not actually be exposed to the challenging items, thus thwarting the public policy goal of increasing exposure to diversity.

More sophisticated presentation techniques, such as aspect browsing in *NewsCube*, may have more potential for showing diverse items to challenge-averse individuals. Agreeable items might be shown on the front page, with challenging items on the same topic linked from the agreeable item page. A similar idea might be based on the presentation used by the political news aggregator *Memorandum*. *Memeorandum* groups items by topics. The front page includes abstracts for top items with links to other items on the same topic below the abstract. To appeal to diversity-seeking individuals, the display might be personalized to show a top-level item and abstract for any topic from a supportive source, with more challenging items appearing in the links. A more complete discussion of the design space appears in the next chapter.

Another limitation of this study is that we have reduced the political spectrum to two broad points of view, and that our subjects only include people whose views fit at the ends of this axis. We do not know if our results generalize to people who are less partisan or more independent. People who are less partisan and people who are not as politically aware (such as any potential subjects excluded or deterred by our requirement for some political knowledge) process political information differently (Zaller 1992), and so our results may not generalize to such individuals. This limitation is another reason that our estimate of the percentage of diversity-seeking individuals must be taken with skepticism.

Individual Preferences

Our finding of individual differences in diversity preferences is one of the more important results of this study. It suggests directions for future research, cautions about the design and interpretation of studies of selective exposure, and may help explain some past contradictory selective exposure results.

Our study is a caution against research designs intended to study the average human's selective exposure preferences. We started out with this intent – to resolve whether people, as a whole, were challenge averse, support-seeking, or diversity-seeking

in their preferences for political news aggregators. An analysis that did not account for individual differences would have produced a fairly strong challenge averse result for challenge aversion; the simple OLS regression model

$$\text{Satisfaction} = \beta_1(\% \text{ agreeable items}) + \beta_2(\% \text{ agreeable items})^2$$

shows a significant effect for percent agreement ($p < 0.01$) but not for (percent agreement)^2 and an adjusted R^2 of 0.40. Though a Breusch-Pagan test would have indicated heteroskedasticity, we might have concluded that people are, overall, challenge averse.

The finding of individual differences may be a partial explanation for the contradictory results of many selective exposure studies. It is part of a growing body of evidence that suggests that diversity preferences vary by individual and by context (e.g., Esterling et al, n.d.; Hinz et al 1997; MacKuen et al 2010). In some contexts, such as the deliberations studied by Esterling et al (n.d.), more people may be diversity-seeking, while in other contexts, such as media choices, more people may be challenge averse. If studies are designed to measure the average human and do not include ways to tease out individual and contextual differences, they will miss the actual distribution of preferences, and instead come to a conclusion that people are, overall, diversity-seeking, challenge-averse, or neither. Future work should continue this exploration of preferences as predicted by individual and contextual attributes, such as by individual-level analysis or studying the preferences of individuals across different tools. Though we find that at least some people are diversity-seeking and at least some people are challenge-averse in their preferences for political news and opinion, we do not know the distribution of preferences in the population as a whole, or if an individual's preferences are common across topic areas, or if someone who prefers to avoid challenging political opinions may seek out challenging opinions about which baseball team will win the pennant this year. Given data about which articles people click on, it might be worthwhile to formulate alternative stopping rules, analogous to those hypothesized in the arena of information acquisition

for decision-making or design, and to estimate their prevalence or the conditions under which people make use of different stopping rules (Browne & Pitts 2004).

Future research also should go beyond short-term measurements and emphasize longitudinal studies. Preferences may vary quite a bit with long-term use of a news aggregator. For example, diversity seekers might prefer diversity one day but tire of it in the long run. Similarly, people who currently are getting diversity might be happy with all agreeable items in the short term, but may not want only supportive items in their day-to-day news source. Other factors, such as whether one's political party is in power, may also affect an individual's diversity preferences over time.

Conclusion

In this study, we find a possible reconciliation of the conflicting theories of diversity-seeking and challenge avoidance: they correctly describe the preferences of different groups of people. Contrary to the implicit assumptions of previous research on selective exposure, neither diversity-seeking nor challenge-avoidance appears to be a fundamental facet of human behavior that describes all people.

The presentation techniques of sorting and highlighting are not very helpful at making challenge more appealing to the challenge-averse people, except that highlighting may make a very small percentage of challenging content palatable. Future work should study additional presentation techniques, including more sophisticated displays of challenging and supporting content. Also, rather than trying to increase the percentage of challenging information in the collections shown to challenge-averse readers, it may be more effective to serve well the needs of those who are diversity seeking and provide them with the means to spread insights they gain from challenging content to the people who avoid such exposure in their everyday news reading.

CHAPTER V

Presentation: *Foresight* and *Hindsight* Widgets[7]

For those seeking to find ways to present diverse views to challenge-averse individuals, the results of the study in Chapter IV are disappointing. Neither presentation technique increased satisfaction with diverse collections. In this chapter, I briefly outline the design space for such widgets and then two field studies for evaluating these widgets. One of the studies is complete: I review the study design and the challenges we faced in executing it. The second study is in development; I discuss its design and why we believe it is superior to the first study in the chapter.

Topic		Question	Findings and outcomes
News aggregators	Presenting diverse collections	• What is the design space for visualizations intended to increase the balance of political items that people read online? • How can / should options within this design space be evaluated?	• A sketch of a design space, including what information can be presented and when. • One completed but flawed experiment, lessons learned, and a new design.

Table V-1 Summary of questions and findings in this chapter.

---

[7] Daniel Zhou was responsible for the article classification portion of this study. Peter Andrews and Brian Ford built the Firefox extension and study website for the *Digg* and *Reddit* study; Jeremy Canfield contributed to its design. Cat Hong Le, Erica Willar, and Alexis Smith helped coordinate human subjects review and award subject incentives. Paul Resnick collaborated on all aspects of the study.

The design space for presenting political opinion diversity

To help design and select which alternative visualizations to test, I outline a design space for interventions to increase the diversity of political news to which people are exposed, or at least make them better aware of their behavior. Features include the time at which the information is communicated and the type of information communicated. Most of the visualization approaches I discuss are designed to take advantage of one of two ways of combatting selective exposure: priming a norm of balance or fairness or increasing curiosity about the different views on a story.

When the information is communicated

A system can present the reader with information about an item or list of items at three times: before a user elects to read items (which I will call *foresight)*, as the user reads items (*"currentsight"*), or after a user reads items, usually in aggregate (*hindsight*).

Based on the principle of *kairos* – intervening at the right time and place (Fogg 2007) – foresight is likely most effective for shaping a user's behavior as it occurs, since the reader can consider the information when they make their choice. By contrast, currentsight may only increase awareness of what they are doing now. This awareness may affect future decisions or their perception of an item's claims, but the choice to read the item has already been made. Finally, hindsight can reveal, to a user, their own reading patterns of which they may not be aware. By doing so, it can increase their awareness and understanding of their behavior over time and prompt reflection (Li et al 2010).

The different presentation times are not mutually exclusive and likely complement each other. For example, a hindsight visualization might make a reader aware of an imbalance in their reading history, and a foresight visualization could then help them identify articles to read that would help to balance out their information diet.

What information is communicated

The design of representations of viewpoint diversity of a reader's information diet also involves many choices about what information to communicate. A visualization can simply describe the user's behavior or choices, can include descriptive social information, or it could go a more prescriptive direction, including norms about what the reader should do, encouraging them to set goals and then monitor their progress against those targets, or by providing recommendations.

DESCRIPTIVE INFORMATION

Visualization might include only descriptive information about what the user could read, is reading, or has read (such the predicted bias of an item or a histogram of the reading behavior). It might also include social information about what the user does compared to others, such as how many and which friends have read an item, how their reading history compares to friends', other visitors to a site, or similar others. This social information could serve as social proof of the value of other stories or reading patterns, or it could simply make people curious about what their friends and people like them are seeing.

NORMATIVE INFORMATION

The visualization could also add explicitly normative information according to progress against a designer's goal or a commitment the user had made. Studies have shown that solely informational messaging about what others do, such as a comparison of one's energy to similar homes, can have a "boomerang effect" on individuals who are performing better than their peers (Schultz et al 2007). Adding a normative message, such as a sad emoticon in reaction to poor-performing individuals and a happy emoticon in reaction to high-performing individuals, can eliminate the boomerang effect for people who are already performing well while still persuading those who are not performing as well (Schultz et al 2007).

People appear to believe in a norm of accessing diverse and balanced views. Fox News, generally regarded as one of the most ideological television news sources in America (Rendall 2001, Kull et al 2003, Pew 2005, Ramsay et al 2010, Krosnick & MacInnis 2010), finds value in marketing itself as "fair and balanced." In a speech at the University of Michigan's 2010 commencement, President Obama implored:

> *Today's 24/7 echo chamber amplifies the most inflammatory sound bites louder and faster than ever before.*
>
> *It's also, however, given us unprecedented choice. Whereas most Americans used to get their news from the same three networks over dinner or a few influential papers on Sunday morning, we now have the option to get our information from any number of blogs or websites or cable news shows. And this can have both a good and bad development for democracy. For if we choose only to expose ourselves to opinions and viewpoints that are in line with our own, studies suggest that we become more polarized, more set in our ways. That will only reinforce and deepen the political divides in this country.*
>
> *But if we choose to actively seek out information that challenges our assumptions and our beliefs, perhaps we can begin to understand where the people who disagree with us are coming from. … if you're somebody who only reads the editorial page of the New York Times, try glancing at the page of the Wall Street Journal, once in a while. If you are a fan of Glenn Beck or Rush Limbaugh, try reading a few columns on the Huffington Post website. It may make your blood boil, your mind might not be changed. But the practice of listening to opposing views is essential for effective citizenship. It is essential for democracy.*

Overall, people reacted positively to this message during and after the speech (Garrett & Resnick 2011, as well as the *Ann Arbor Chronicle's* notation and my own cursory examination of reactions online). This is possibly because most people think the other side(s) should listen to their own perspective more, but also possibly because most people agree with the norm even if they do not choose it in the moment or are not aware of their own behavior. If the latter explanation is true, tools that remind people of a norm of fairness and that highlight when they are behaving inconsistently with this norm may be particularly effective.

The choice to add normative information about progress relative to a designer- or user-supplied goal implies a design choice about goals, though this does go beyond a visualization feature. Does the feature present a goal, and if so, should the goal be user-set, set by social information (what others do), or set by the system's designer? If the feature promotes setting and/or pursuit of a goal, goal-setting literature provides a number of techniques for making it more effective. In general, goals should be specific and high, yet achievable, and people should have a way of monitoring their progress toward them (Seijts et al 2004, Locke & Latham 2005). The individual should be committed to the goal, and techniques such as asking them to record or state their commitment or explaining why the goal is important can increase their commitment (Locke & Latham 2002, Cialdini 2009).

Systems can also present recommendations from a variety of sources. Here, too, there are many options. Stories on a particular topic could be clustered, such as in *NewsCube*'s presentation (Park et al 2009), possibly increasing readers' curiosity about the different clusters. Many news sites now recommend stories that are popular, both among their general audience and among a reader's friends. If the user sets a goal, the system could also show recommended items that would help a reader meet his or her goal, such as conservative items to balance out a liberally slanted reading history.

First study design

In the first study, we sought to test two types of visualizations – a foresight widget showing articles' predicted bias and a hindsight widget indicating reading history – as well as their interaction. As with other studies in this dissertation, we focused on news aggregators where the main content is selected by users' votes and that have substantial politics sections, specifically *Digg* and *Reddit*.

We created a user-installable extension to the Firefox web browser that augmented *Digg* and *Reddit* with two widgets: one that annotated individual articles according to their predicted slant and one that showed the reader a representation of their reading history. The extension was deployed in a field experiment, first to *Digg* users and then to both *Digg* and *Reddit* users. Our outcome variable was the political distribution and quantity of articles people read.

Classification of articles

Like the previous study, this experiment required articles with a known (or at least, reliably predicted) political slant. For *Digg* articles, we could obtain fairly reliable classifications by classifying some articles' bias, classifying the *Digg* users who voted for those articles (who could be assumed to share political positions with the item), and then using those user and item classifications to classify additional articles. This method is detailed in Zhou and Resnick 2011. In addition to this automated method, we also used human raters to classify all of the articles. This allowed us to generate an immediate, automated classification for many items, and a more accurate, but slower classification for all items.

We classified articles that appeared in the *Digg* politics section, which we crawled every ten minutes. Whether or not we could automatically classify the article as liberal, conservative, or neither, all articles were sent to a panel of workers on Amazon Mechanical Turk to be classified by human raters. When the human raters disagreed with the automatic classification, we used the human rating.

For reasons I elaborate on later, we chose to expand the study to include the *Reddit* site, items, and readers. Unlike on *Digg*, the majority of *Reddit* users' votes are not accessible through the API. This made classification of *Reddit* items slower because we could not use an automated process to generate an initial classification, as we did with *Digg* items. On *Reddit,* we monitored and classified articles from a number of political

sections (or "subreddits"), including: /politics, /uspolitics, /democrats, /republican, /conservative, /progressive, and /libertarian. The sections were selected for being the most popular sections with primarily political content. For articles that appeared only on *Reddit,* an initial, automated classification was not available unless the item had already appeared on *Digg.*

Because the human classification process took time, and because the classification from Mechanical Turkers sometimes differed from the initial classification, users sometimes saw articles before they were classified or with a classification that differed from their final classification. In these cases, we tracked their click behavior according to what they saw at the time for the purposes of our analysis, but updated their reading history and our hindsight widget (described below) according to the article's eventual classification.

The BALANCE widgets

We designed two widgets: one that highlighted political articles according to their predicted slant or bias (foresight) and one that reflected users' reading behavior for the week to date along with normative messaging (hindsight).

FORESIGHT

The foresight widget (Figure V-1) replaced the standard vote box on *Digg* or *Reddit* for political articles, shading the background (normally yellow) according to the predicted bias: blue for liberal, red for conservative, and grey for independent or neutral. This widget would help make people aware of the predicted bias of articles as, or before, they chose to click on and read them. This widget is similar to, and partly inspired by, the *Memeorandum Colors* widget,[8] developed by Andy Baio. *Memeorandum* clusters current

---

[8] http://waxy.org/2008/10/memeorandum_colors/

news stories with articles from a variety of sources; *Memeorandum Colors* highlights those source links according to the predicted bias of the source.
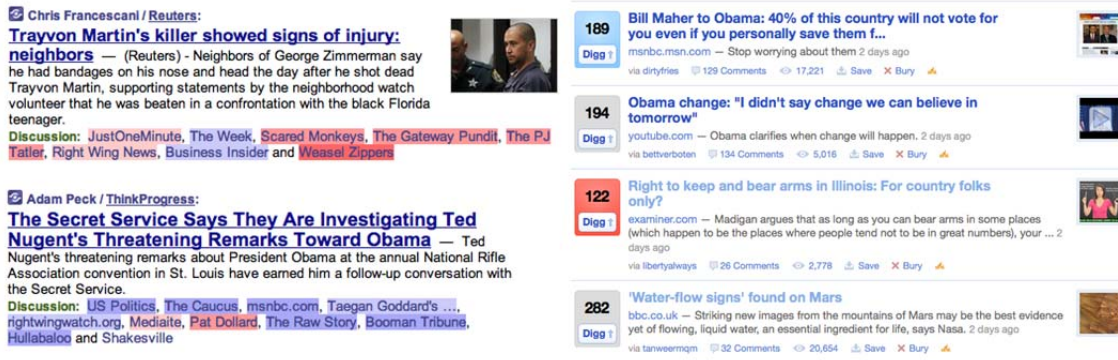
Figure V-1 Example foresight visualizations. Left: *Memeorandum Colors*. Right: our foresight widget on *Digg*.

HINDSIGHT

The hindsight widget showed an approximate histogram of the user's liberal and conservative articles clicked on (which we used as a proxy for read) for the week to date. This was motivated by the idea that while many people might agree, in principle, with the normative goal of reading diverse news, they might not realize just how skewed their own reading behavior actually is. We considered several possible designs for this widget (Figure V-2).
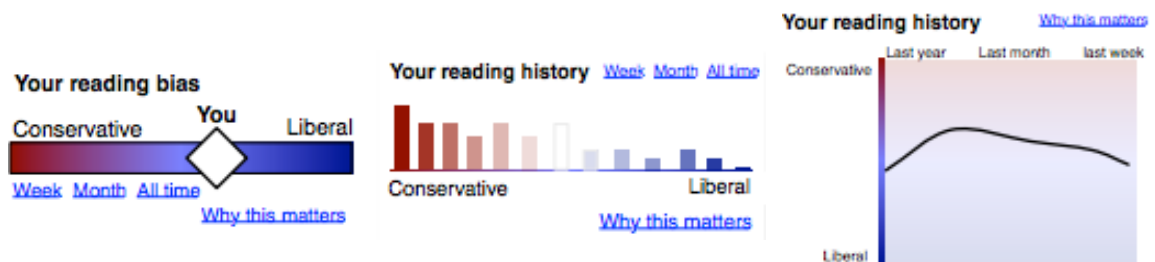
Figure V-2 Early hindsight widget mockups.

As noted earlier, communicating a norm can increase the persuasive power of information, particularly when an individual's present behavior deviates from that norm (Schultz et al 2007). A simple histogram of liberal and conservative items does not communicate any form of norm – just the individual's own behavior. We believed that

communicating an injunctive norm (that is, what one *should* do (Cialdini et al 1991))

might be more effective than a simple histogram. Apropos to the name of the research

project – BALANCE – we chose to implement the hindsight widget as a character on a

tightrope, with his balance affected by the histogram. If one's reading behavior is too

skewed, the character is in peril of falling from the tightrope (Figure V-3). We hoped that

this would encourage subjects to consider the norm of fairness and balance, one of the

methods for counteracting selective exposure (Sears 1965), though it was also possible

that they would treat the exercise of balancing the character as a bit of a game. The

hindsight widget reflected history for the week-to-date, which was reset on Sundays.

Figure V-3 Example hindsight visualizations. Left: No articles (beginning of week), center: unbalanced, right: balanced.

The combination of the foresight and hindsight visualizations allows people to see

if their reading behavior is imbalanced and then skim a list of articles to find some

balancing selections to immediately take corrective action. In contrast, we hypothesized

that the widgets might have different effects when seen alone. Readers viewing only the

foresight widget might use it, intentionally or unintentionally, to identify and read

agreeable articles. Readers viewing only the hindsight widget might try to balance their

reading history, but have a more difficult time doing so without the foresight widget to

help them quickly identify the articles needed to balance their reading, and thus a

hindsight-only view might be less effective than hindsight plus foresight.

Experiment design

To test the effects of our widgets on news aggregator users' reading behaviors, we designed a field experiment with a within-subjects design. After clicking on one of our ads or links, each subject would land on the study website, where the project was explained. If interested, they would then:

1. complete the consent process,

2. create an account by providing their email address, creating a password, and providing their mailing address (for delivery of participation incentives, described below),

3. complete a survey, consisting of items from the Big Five personality inventory (Goldberg 1990)[9] and two seven-point scales for political preferences (extremely liberal to extremely conservative and strong Democrat to strong Republican),

4. and download and install the Firefox extension. This extension was custom built for the logged-in user, so that we could associate their survey responses with the behavior recorded by the extension.

Once installed, the browser extension would annotate their news aggregator (at first *Digg*, and later *Digg and Reddit*) with the widgets according the subject's current experimental stage. Each stage would last until the user had clicked through to a certain number of classified articles during that stage. The stages were foresight only (30 articles), foresight + hindsight I (60 articles), hindsight only (30 articles), no widgets (30 articles), foresight + hindsight II (duration of use / study). During the stages without a hindsight widget, a placeholder with information about the study was shown. This combination of stages would allow us to measure the effects of each widget on its own and of the widgets together. By returning the subjects to a no-widget condition for 30 articles, we would also be able to learn whether any behavior change was (at least briefly) lasting without the

---

[9] We used 26 questions, with a focus on openness: ten openness questions and four questions from each of the other dimensions of extraversion, agreeableness, conscientiousness, and neuroticism.

support of the widgets. It would, however, risk that one stage influenced the next, and we did not counterbalance the stages to control for this possibility.

We originally intended to display the widgets and track reading behavior on just the *Digg* site, because we could quickly generate fairly reliable, automatic classifications for *Digg* articles and because developing and maintaining a plugin for one site required less developer time than doing so for two sites. Difficulties in advertising and recruiting Digg readers to be subjects led us to later expand to include *Reddit* readers and items as well.

Subject recruitment

We planned to recruit subjects through advertisements on *Digg*. To incentivize participation, active subjects (i.e., those from whom the widget logged activity on any given day) would be eligible for a daily drawing for *Digg* (and later *Reddit*) merchandise, such as hats, shirts, stickers, laptop sleeves, and mouse pads.

After delays in arranging to run the advertisements on *Digg*, we began recruiting through word of mouth (posts to mailing lists and social websites) and advertising on the Google AdWords and Facebook networks for keywords such as politics, political news, and political opinion, beginning on 5 August 2011. Less than twenty people signed up in response to this initial campaign, so we began modifying our extension, classification processes, and study design to work on the *Reddit* site as well, and were able to deploy a revised version that worked on both *Digg* and *Reddit* on 16 August.

Eventually we (1) expanded the study to include *Reddit,* and (2) were able to run our recruitment advertisements on the *Digg* (starting 20 August) and *Reddit (*starting 18 August) politics sections. This resulted in a major traffic increase to our study website and a large increase in signups. 337 people consented to participate the study. 11 people subsequently abandoned it at the account creation stage, 11 more dropped at the survey, and 62 abandoned the study when they did not install the Firefox extension. 253 subjects completed the signup and installation process.

Results: major issues with study design

By October 28, only 33 of those 253 subjects had moved past the foresight only stage: seventeen reached the foresight + hindsight I stage, five reached the hindsight only stage, four reached the no-widgets stage, and six reached the end of the study (foresight + hindsight II). We only ever received data from 178 of the 253 subjects (70%), suggesting that 30% either installed the extension on a browser that they never used, that they were not actually *Digg* or *Reddit* users, or that there were undiagnosed compatibility issues between our extension and others installed on the computer. These numbers are too small to use to make meaningful conclusions about the widgets' efficacy at an overall level, let alone at levels of individual personality differences. What went wrong?

First, the click through rate on political articles was very small. Subjects read political articles at a rate far lower than we expected. For the 178 readers for whom we received data, they visited *Digg* and/or *Reddit* pages listing stories an average of 5.88 times per day (median=1.89, *sd*=11.46). They clicked through to 0.95 (median=0.40, *sd*=1.80) political stories per day while they were in the study. A few outliers inflate both of these averages (Figure V-4). Perhaps skimming article abstracts on *Digg* or *Reddit* is sufficient news exposure for users of those sites, and so they rarely click through to news stories. It is also possible that our subjects were getting most of their political news elsewhere – directly from news websites, through other aggregators such as *Google News*, through mobile applications, and so on – and were getting something else, such as a social experience[10] or news on a different topic, from *Digg* and *Reddit*. Or perhaps people mostly just glance at headlines and snippets, regardless of the site. Whatever the cause, there are two severe problems with the combination of this reading pattern and our study and intervention design. First, the median user would take more than two months to progress

---

[10] Though if so, this social experience is not on the comment pages of the political articles: they visited only 0.19 political story comment pages per active day on the study.

through just one stage of our study. Second, the intervention was designed to reflect weekly political reading history; for the median subject, that history would have been based on fewer than seven articles.
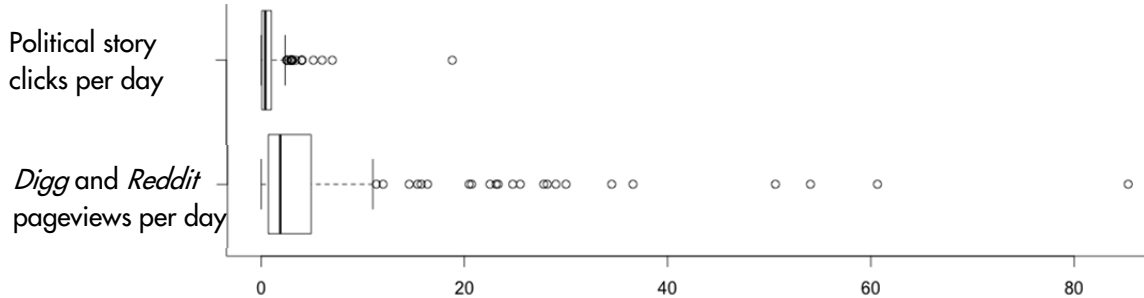
Figure V-4 Distribution of average political story clicks and *Digg* and *Reddit* pageviews per day for each subject. Boxplots with Tukey method.

Recruitment was our second challenge: had we been able to recruit more subjects, we would have at least been able to draw conclusions about those who read many articles. *Digg* and *Reddit* actually account for a fairly small amount of the Internet news audience, with roughly one-tenth the daily visitors as sites such as *Google News* (Purcell 2010). We believed that recruiting by advertising specifically to the sites' visitors would help us reach people who would be a good match for the study. We were also challenged by declining Firefox market share between our study's inception and launch; two-thirds of the visitors to the study website were using other browsers. This was reflected in comments about our study on *Reddit* as well as emailed feedback: many wished for and requested a Chrome extension. If many users installed our extension in Firefox and then switched right back to using Chrome or another browser, this might explain why we never received visit data from 30% of the users who installed the extension.

Finally, even if reading behavior and audience were different, the article pool on *Digg* and *Reddit*'s politics sections was and is limited. During the study, we classified a total of 33,358 articles, 18,606 on *Reddit* and 14,752 on *Digg*. On *Reddit,* 52% of the classified articles were liberal, 34% were neither or independent, and only 13% were conservative On *Digg,* 46% were liberal, 32% were neither or independent, and 25% were

81

conservative. From our anecdotal observations, the home pages were often more skewed than the politics sections: the minority rarely made it to the front page, though our logging data is unfortunately not in a format that allows us to quantify this. Thus, a reader with a liberally skewed reading history had very few conservative articles that they could read to balance out their history.

One other difficulty arose unrelated to the study design. Around the time that we expanded the study to include *Reddit* users, the pre-survey stopped recording responses to the Big Five personality questionnaire. We did not observe this problem until after the study, at which point it was too late to contact subjects and ask them to re-take the survey, or to fix if for future subjects. The occurrence of this issue and our delay in noticing it teach two important lessons: (1) use unit tests to validate that all existing code functions before deploying updates to study software, and (2) build a more comprehensive dashboard to enable study administrators to quickly note any data problems that do occur.

The sum of these difficulties with subjects' limited reading of articles by clicking through links on *Digg* and *Reddit*, recruiting subjects, offering a sufficiently diverse selection of political, and technical issues caused this experiment to fail. The challenges, however, can be at least partially mitigated with a different study design that is less reliant on *Digg* and *Reddit*.

An alternative study design

Ideally, an improved study design would allow us to observe subjects' reading behavior across all of their political news sources, rather than just articles they arrive at through *Digg* and *Reddit,* be able to classify all of these observed items, and offer the user sufficient value that they would want to install it and participate in the study, including answering a questionnaire and sharing their information with us.

Observing behavior

There are a number of possible technical implementations that would get at least part way there, including a system-wide proxy, an application that analyzes and transmits a user's history once, and a browser extension that accesses a user's entire web history, both before installation and on an ongoing basis.

A system-wide proxy. In this design, the user would configure all web traffic on their device or devices to pass through a server administered by researchers. This server would record the traffic that passes through it. This would allow us to track behavior from all web browsers and applications (unless they overrode the proxy), across multiple devices, including mobile. Workplace computers would likely still be excepted, since many companies use their own proxies and/or do not give their employees the administrative rights required to set a proxy. Unfortunately, it is also somewhat resource-intensive to run a proxy and requires high uptime in order to keep subjects happy. Menchen-Trevino and Karr pursued exactly this approach with their *Roxy* software, but note that renting a dedicated server was necessary in order to support their 46 study participants (2011). Their proxy captured the entire request and content for all sites visited, though the participant could select a private session or a guest session.

History analysis application. Another approach would be to write an application that users download and run to analyze their entire history. The analysis could be conducted locally (preserving more privacy), with results submitted our system, or remotely on our system. Such an application could read from the history directories or files for all major web browsers. Variations would need to be written for different operating systems, and it might fail on non-standard configurations.

Browser extension(s). Similar to the original study design, we could also use a browser extension that monitors a user's tabs (collecting history as they use their browser) and/or accesses and classifies their entire browsing history. As with the application,

classification could occur locally or remotely. Unlike the application method, a single extension could be cross-operating system, and users may also be more willing to install a browser extension, with limited permissions, than a full application. Collating history for the same user across multiple browsers would require that they create a log-on and that they install multiple extensions. Neither the browser extension nor the history analysis application would be able to observe traffic from mobile applications or non-browser desktop applications (such as an RSS reader), and, unless the user syncs their browsing history between their mobile and desktop browsers, would not be able to track articles read on the mobile browser.

Classifying articles

There are two primary questions for the classification of articles: first, what classification method should be used, and second, should classification occur server-side or client-side?

When we had only a limited set of articles – the popular political articles on *Digg* and *Reddit* – it was possible to use a panel of Mechanical Turk workers to classify all articles in a reasonably short period of time. We paid at least $0.02 for each rating, with four ratings per item; this approach is obviously not suitable for classifying entire web histories for hundreds or even thousands of users.

Fortunately, a variety of researchers have succeeded at using machine learning to classify articles as liberal or conservative, fairly reliably, over time. This includes work by Oh et al (2009) and Dehghani et al (2011). These classifiers would need up-to-date training data to reflect current political figures, positions, and topics, but classifiers using *Digg* articles and votes, or articles linked from classified blogs, similar to the approach used in Chapters II-IV and further developed by Zhou, Resnick, and Mei (2011), could provide a steady source of classified articles to use as training data. These classification methods, however, require accessing the website, downloading the page, extracting the article, and then classifying it first as political or not, and, if political, then according to its

bias. Thus, while they would be appropriate for classifying items as a user accesses them,[11] they would not be suitable for providing a foresight visualization, and, if used to process a web user's existing history, there would be a delay before feedback is available.

An alternative approach is to classify simply based on an item's URL. This is how *Memeorandum Colors* provides its foresight visualization: it indicates the political leaning of sites that typically link to the source, though the individual item's bias may differ. There is also precedent for building hindsight widgets based on URL-based classification. In 2010, *Slate* released a tool that let people check the bias of their reading history against 112 websites that had been classified by Gentzkow and Shapiro (2011) and then showed them their isolation score and how they compared to other *Slate* readers who had used the tool (Matlin et al 2010).

Enticement to participate

There are different ways we can motivate people to participate in the study. The first is the standard approach of incentives, such as a flat incentive for participation in the form of money, a gift card, or a token of appreciation, or prizes or a lottery, such as used in the study we did try. Alternatively, we could offer a tool that is sufficiently appealing on its own. Both *Memeorandum Colors* and *Slate*'s isolation index tool attracted substantial coverage from traditional media sources and bloggers, showing that this strategy can work and that many people find such tools interesting. On the other hand, our experience with recruiting for the previous study demonstrates that generating interest and coverage sufficient to get these signups can be quite difficult. At a minimum, we have to build an interesting tool and get it in front of the right people. This is probably harder for some types of visualizations than others. In aggregate, such as in a hindsight visualization, URL-based classification may provide acceptable accuracy, but even a modest number of

---

[11] Article extraction could occur either in the browser (through the extension) or through the proxy, depending on which method is being used to observe the user's web activity.

inaccuracies might cause users to become frustrated with a foresight visualization – which exposes the classification of each individual item – and uninstall our extension.

By offering incentives, we would likely get a sample that is somewhat more representative of the population – especially if care is taken in recruitment. By offering only a compelling application, though, we could attract participants who are more likely to represent the actual users of such a tool, and who are more likely to explore its features and make use of it.

Proposed study design

There are many possible studies within this design space, but I propose one that uses a browser extension, URL-based classification on our server, and a hindsight widget. The browser extension allows for collection of past and ongoing data and requires little user configuration. To reduce barriers to installation, the extension will generate its own unique identifier when installed. There will be no signup or account creation process. URL-based classification is most suitable for real-time classification of a large number of sources, and should provide suitable accuracy, at least in aggregate.

Subjects would be recruited based on their interest in seeing the hindsight information, though an additional panel meant to better represent the general population of Internet-using Americans could be recruited for comparison. To classify the largest number of sites possible, URL-based classification would be based on the *Memeorandum Colors* classification as well as the blog and *Digg*-vote datasets outlined in Chapters II-IV. A rough prototype of the Chrome extension, which shows a persistent indicator of the subject's balance in the browser window and provides more detail when clicked, is shown in Figure V-5. The persistent icon helps people to monitor their behavior, with little effort, and to make changes in response. Participants have responded positively to using persistent, glanceable visualizations for self-monitoring in studies in other domains, such

as physical health (Consolvo et al 2008) or personal transportation choices (Froehlich et al 2009).

Figure V-5 Example Chrome extension with hindsight visualization. The icon in the browser provides persistent indication of the reader's history for the week, while clicking shows a larger view.

After the user installs the extension, the extension will transmit their previous month's web history to our server. Before being able to see our analysis of their browsing history, the subject would be prompted to complete a brief questionnaire, consisting of a twenty-item version of the Big Five personality index (Donnellan et al 2006), age, gender, and two questions about their political preferences (the same two as in the last study). After showing a representation of their reading balance, the extension would continue to run as they browse, updating the visualization and transmitting their browsing history to our server.

This design allows us to collect and retain users' browsing history, including their history before they installed the extension and their browsing activity after installation – either their entire history or just the subset of URLs that matches our list of news and opinion sites. This allows further analysis beyond what we are able to capture in real time. Privacy is a concern in this design, but this concern is mitigated by: (a) transmitting the data over a secure connection, (b) not collecting the subject's email address or identifiable information as part of the registration process, (c) possibly discarding some particularly sensitive URLs (e.g., social network sites, search terms) on the client side before

transmission or limiting collection to a whitelist of URLs of news sites, and (d) disabling IP-logging on the web server. The data we would collect is already collected by many companies (whether or not users know it): Google receives all information typed into Chrome's "omnibox" (Chrome's combined address and search bar), and *Twitter* and *Facebook* are able to track users across sites that use their tweet, like, and share widgets, unless those users have disabled third-party cookies.

PERSONALITY AND BROWSING HISTORY

Even without an intervention or experimental groups, the combination of subjects' browsing history, their personality survey responses, and their political leanings can be used to answer several questions about how people access news online. This includes analyzing for individual differences, based on personality attributes, on the number of news articles read, their political balance, and possibly ratio of headlines they have seen to the number of articles they clicked on to view in detail. Political balance would be represented either as their KL divergence from an ideal, balanced reading (50% conservative, 50% liberal) or from the balance of liberal, neutral, and conservative sources present in our URL classifier, weighted according to their audience. We have previously conjectured that people who score high on openness might read more articles and have more balance in the articles they read, for example. Analysis would be regressions using the five personality attributes, gender, age, and political leaning (a combined variable) as the independent variables and their reading behaviors (described above) as dependent variables. We would also be able to reproduce Gentzkow and Shapiro's isolation index analysis (2011), with the addition of the personality data.

One limitation is that we are only able to describe their desktop browsing history in one web browser. If, for example, someone who is more open also splits their web browsing and news readership across multiple devices and web browsers, we may not observe the hypothesized additional reading (or may even observe less reading) even if it

is occurring. This is one way in which the study design could made be stronger by using a proxy to collect history rather than a single browser extension, but the added complications of recruiting willing subjects, helping them set up the proxy on all of their devices (with many possible operating systems), and maintaining a proxy with sufficient speed and uptime to retain subjects make a proxy impractical.

TREATMENT GROUPS

Of the design space of visualizations, the balance-man hindsight visualization – an icon persistently indicating the balance of a user's reading history, as well as an injunctive norm about what that balance should be – is a promising starting point for investigation. This leaves open the question of what to use for the comparison group. One, easy option is to compare their history post-intervention to their history pre-intervention, in a within-subject design. This presents some challenge for analysis. Some subjects will be imbalanced in their recent history, but, even without the intervention would regress to the mean during the study period. If, however, we do not see other subjects becoming more imbalanced in their browsing history during the study period, we would still able to attribute observed effects to the intervention. This design, however, also conflates the effect of being in a study (and being prompted to attend to one's reading balance) with any effect of the widget. It is also vulnerable to history effects: if a major news story (e.g., one that people cannot ignore) breaks during one of the time periods and is covered more by one side or another, it will affect the balance of subjects' reading. The last two issues are more problematic confounds.

One way to control for this would be to deliver different strengths of the history visualization. The first and most powerful would be the persistent icon, updated as subjects' browse the web and reset weekly. The second would be a weaker version, which displays a report when they first install the widget and then does not update again until a week or month later. This would make the experiment into a test of the importance of a

persistent and regularly updated self-monitoring tool versus one that is infrequently updated. Alternatively, we could deploy the man-on-tightrope visualization to some subjects and a simple histogram (in which no stick figure is in peril) to others. Again, this would be a different experiment on the effect of the different styles of hindsight visualization and our approach of communicating an injunctive message using the stick figure.

Another option is to use a foresight-like visualization as the control. I have been reluctant to deploy such a tool as part of this study, because URL-based classification will be less reliable for individual items than in aggregate. Visible mistakes in the classification of items could cause people to abandon the tool or to not trust the aggregate ratings, and thus diminish its efficacy for encouraging them to reflect on and possibly change their reading behavior. The positive reception to *Memeorandum Colors*, however, suggest that a foresight tool using URL-based classification is viable, at least for one news aggregator and its limited set of sources. We could use a "currentsight" visualization as the control by replacing the BALANCE-man icon with a single square indicating the predicted bias of the user's current page (rather than shading links, as our previous foresight widget and *Memeorandum Colors* did). It is also possible that such a tool could be used to collect ratings for additional individual pages and sites, simultaneously giving readers a way to take action, other than uninstalling the extension, about frustrating misclassifications and improving our ability to classify pages and sites for future studies. This widget might also change subjects' behavior somewhat, but, because it only shows classifications for the item they are currently viewing (i.e., items they already have chosen to visit), the effect should be small.

Finally, we could use an entirely different tool for the control. One possibility is the *CubeThat* system that Sidharth Chhabra is developing (Figure V-6). *CubeThat* collects and then clusters news stories according to the approach developed by Park et al for *NewsCube*

90

(2009). When a reader visits an item for which there are *CubeThat* recommendations, they

can click the icon in the browser to view clusters of links to recommended articles on

different facets of the story. Park et al demonstrated that such clusters increase readers'

exploration of news items, possibly by making readers curious about the different aspects

or the differences between clusters while also not indicating that some articles might be

from aversive sources. At the time of writing, *CubeThat* is nearly ready for deployment.

Subjects could download and install an extension that provides either the BALANCE

hindsight widget or *CubeThat*, and we could then compare the reading behaviors of the

two groups.

Figure V-6 *CubeThat* running on a news article on the CNN website. Credit: Sidharth Chhabra.

There are two primary issues with using *CubeThat* or another tool for the

intervention. First, we do not think that *CubeThat* is a placebo; instead we believe that it

should increase subjects' news exploration. If we compare two interventions, both of

which we believe should work, we may not be able to tell if the widgets have no effect or if

both have a similar effect (though this is mitigated by having subjects' pre-intervention

web histories). It would, more accurately, be a test of two ways of combating selective

exposure: priming a balance norm, with a hint of gamification, using hindsight versus

increasing readers' curiosity about other aspects to a story using *CubeThat*. Second, marketing this study to subjects, without using incentives, could be a challenge, since we would need to randomize which extension the subject received. People could be reluctant to sign up for a study and install an extension that will randomly deliver one of two very different experiences.

In contrast, I can advertise an extension that shows the hindsight widget, either after a delay or immediately, using the same, simple recruitment materials. Thus, I prefer a traditional control: they install the extension, it reports their history to us, but they do not get to view their history until one month later. There is some risk, though, that if subjects do not get immediate gratification, or are asked to wait for a long period of time, they will uninstall the extension. I may, then, decide to also deploy a currentsight visualization as control, and to also use that widget to collect improved website classifications from subjects in that group. At the end of a month, subjects would also be able to view the hindsight widget. A currentsight visualization could also be advertised with similar language, as it presents similar information as the hindsight visualization, just over a different time frame (current page vs. the last week or month). The use of two control groups splits the subject pool into more than one control group, which increases the risk of a Type II error if we are only able to recruit small numbers of subjects. I believe that the potential benefits for recruitment and retention may make it worth this risk.

ANALYSIS

The study design consists of both a between subjects and a within subjects component. This allows for an analysis that compares the differences between subjects' pre- and post-intervention reading behavior across the treatment groups. It would otherwise parallel the analysis described in the personality and browsing history section.

Expected Outcomes

From the revised study, we will be able to learn (1) whether any of the big five personality indicators and/or any of the demographic information we collect predict balance or quantity of political articles read online and (2) the effects one particular hindsight widget on subjects' reading history. If this new study is successful, it will also be a proof of concept for future studies of visualization techniques intended to increase the balance of political articles that people read.

CHAPTER VI

The Prevalence of Political Discourse in Non-Political Blogs[12]

In the previous four chapters, I focused on design challenges and alternatives for news aggregators that select and present political content, as well as individuals' preferences for the opinions that appear on such sites. Many other recent studies of politics in online spaces have also focused on political spaces: Usenet groups (e.g., Kelly et al 2005), blogs (e.g., Adamic & Glance 2005; McKenna & Pole 2007, Koop & Jansen 2009, Yano & Smith 2010), media and news sites and their audiences (e.g., Park et al 2009, Park et al 2011), and political and media accounts on social network sites (e.g., Golbeck & Hansen 2011). These political spaces are no doubt important, but this focus neglects that a good deal of political opinion formation occurs *outside* of explicitly political venues. Even people very interested in politics may be exposed to political news and opinion in spaces not devoted to those topics, and people who are less interested in politics may never visit explicitly political sites. In this chapter, I use a sample of blogs hosted by Blogger.com to show why the studies of online political discussion should be broadened to include non-political spaces.

Some of the most alarming support for concerns about polarization in online political discussions has come from studies of political blogs. For example, Adamic and

---

Glance (2005) found that political blogs rarely link to blogs expressing opposing views. Readership patterns are also polarized: political blog readers are more polarized than people who get their news through television (Lawrence et al 2010). Gilbert et al (2009) analyzed comment threads on several prominent blogs and found that political blogs are echo chambers; a comment was more than three and a half times as likely to agree with the original blog post as it was to disagree.

As previously discussed, other work challenges these alarms. Kelly et al found diversity in discussions within some political USENET groups (2005), and Stromer-Galley found that participants in online political discussions report seeking out diverse opinions, though she was unable to determine the actual diversity in the discussions (2003). Survey research from the 2004 election suggested that readers of online political content are not using the tailorability of the web to filter out contradictory viewpoints and may in fact see a wider range of opinion than counterparts receiving news from traditional sources (Horrigan et al 2004). In the study in chapter IV, I found mixed results: some people actively seek out opinion diversity in their information sources, while many others are averse to politically challenging material.

Empirical studies indicate that online political discussions have the potential to approach deliberative ideals. Price and Cappella created a political chat room to use as a research setting (2002). They introduced a random sample of people to this research setting and then measured indicators of quality of online discussion and its impacts on participants, including opinion change, opinion quality, electoral engagement, social trust, community engagement, distortions, and alienating effects. The researchers observed positive outcomes in discourse quality and civic engagement. Might some of the elements seen as necessary for such deliberations be present in blogs?

Topic	Question	Findings and outcomes
Non-political spaces	• Does political discussion occur in non-political spaces online?	• On *Blogger*, half of the political discussion occurs on non-political blogs.
	• If so, do people treat it as taboo?	• When politics comes up on non-political blogs, it elicits as many comments as non-political posts. These comments engage the political content and include some disagreement.

Table VI-1 Summary of questions and findings in this chapter.

Politics in non-political spaces

Several decades of work within political sociology has shown that political opinion may be frequently, even primarily, shaped through non-political intermediaries, such as friends and family members, opinion leaders, etc. (Katz 1957). This happens in the broad context of social interactions at times and in places where participants are not explicitly seeking out political information (Putnam 2000, Habermas 1962) but where "chance encounters" (Sunstein 2001) with opposing views may occur.

Some recent work has shown the value and importance of studying politics online wherever it occurs, even outside of political and news spaces. Researchers found a correlation between the quantity and sentiment of Twitter mentions of candidates on the one hand and both political debate performance (Diakopoulos & Shamma 2010) and election results (Tumasjan et al 2010) on the other. During the 2010 US midterm elections, 8% of online adults reported posting political content to Twitter or another social network site, and 11% said they discovered on a social network site for whom their friends voted (Smith 2011).

When politics comes up in non-political spaces online, we might expect the discussion to more closely approximate deliberative ideals than conversations in political spaces. The reason is that participants with more diverse views may be present, given that the audience formed around some other topic, and the desire to maintain relationships formed for other reasons may make them more disposed to listen to each other and to

make the effort to frame arguments in a way that opponents will understand. If political discussion does occur frequently in non-political spaces online, then, we argue that such spaces will be important settings for study of online political discussion.

Through survey research, Wojcieszak and Mutz (2009) found that political discussion does occur in non-political spaces. People reported that of online apolitical chat rooms and message boards they were part of, between 30% and 70% "ever talked about political topics or controversial public issues." It is not clear, however, how frequently these topics come up or the nature of the discussions. Goel et al (2010) found that Facebook users are often unaware of differences of opinion with their Facebook friends, suggesting that political topics come up infrequently or that people do not reveal their opinions when they do come up.

Eliasoph's investigations of political speech (1998) found that both jokes and serious discussion about politics in social clubs tended to be met with silence rather than provoke a discussion on the topic, either out of ignorance or to avoid expressing disagreement in a social setting where they did not know the opinions of others or knew that others disagreed. Mutz and Martin (2001) also note a tendency to avoid political disagreement in interpersonal relationships in order to promote social harmony, and Noelle-Neumann (1993) found that this is particularly true in social contexts with diverse views. Given the mixed social contexts of many personal or non-political blogs – where readers may be coworkers, friends, family, acquaintances, potential employers, or strangers – might bloggers choose to stay quiet rather than risking offense?

We are left, then, with important empirical questions. In this chapter, I investigate the empirical prevalence of political posts in different types of blogs, and the reactions those posts get. In particular, I answer the following questions:

- How prevalent are political blog posts on non-political blogs?
- Relatedly, what is the distribution of political blog posts across different categories of blogs?

When readers of non-political blogs encounter political posts, do they treat them as taboo, or do they engage with the political content of the post?

Methods

To study these questions, we used a collection of posts and comments from blogs hosted by the blogging service *Blogger*. Each blog was categorized into one of eight categories and each post was categorized as political or nonpolitical.

Data set and collection

From 6-20 January 2008, we automatically monitored Blogger.com's "recently updated" list, checking it at periodic intervals to identify blogs that were written in English, had existed at least since 31 August 2007, and had at least five posts total. This produced an initial sample of 23,904 blogs written in English and that had a reasonable level of author activity and some longevity. On 3 June 2008, we used the Google Data API to download all of the posts for each of these blogs that still existed. This sampling method does introduce some bias into our sample, as blogs with more frequent posts are favored for inclusion. We wanted to ensure that the blogs in the study had some minimum level of activity and audience, so we further constrained our sample to exclude blogs with less than twenty posts and less than five comments total. The combination of requiring comments and that the blogs existed from August 2007 until at least June 2008 also had the effect of eliminating many spam blogs.

Classification of posts

Human coders classified 6,691 posts as political or non-political. These posts were selected through a combination of purposeful and random sampling. Initially, we drew 2000 random posts from the full sample of posts. While coding these posts as political and nonpolitical, we realized that our set of blogs still contained many blogs that were either spam or not written entirely in English, so researchers looked at each blog and removed

many non-English and spam blogs, reducing the number of blogs to 8,861 (and the number of sampled posts to 1,691). To increase the number of political posts for training purposes, we coded the originally sampled posts as political or non-political, then identified blogs that had at least one political post in the original sample and drew an additional 4,000 posts at random from those blogs. We also added another sample of 1,000 posts drawn randomly from all of the posts in our sample.

We considered posts about public policy, campaigns, and elected or appointed officials as political, and did not restrict this definition to politics in the United States. Posts were classified as political even if the political content was only a brief mention in a much broader post; that is, we coded for the presence or absence of political remarks, not for the primary topic of the post. Comments (if any) were not included in the text to be classified, so the label was based only on what the post author wrote in the post title, keywords, and body. To assess the validity of this measure, two human ratings (by a researcher and an undergraduate student employee) were collected on a randomly selected subset of 500 of these posts, after the two raters first discussed 25 posts that one researcher rater found to be difficult to classify. The kappa score (Cohen 1960) between the two human raters on the 500 randomly selected posts was 0.969, and so a single rating of the remaining posts was collected. The classification yielded 1,676 political posts and 5,015 nonpolitical posts.

The political posts on nonpolitical blogs took a variety of forms. Some posts encouraged readers to vote. Others asked questions (e.g., a lengthy post discussing a child's illness asks "why can't we get universal healthcare for children?"). Some of these political posts were re-posts of something the author received by email or found on another blog. Others included quick, throwaway references (e.g. a quick complaint about a political figure in the middle of a mostly unrelated post). Some of these posts were also much longer pieces, expressing disapproval about a political figure or action, or talking through their decision about how to vote in an election.

With the human-coded data as a training set, we used Weka (Witten & Frank 2005) to classify each post as political or nonpolitical. For tokens, we lowercased each post and took the 10,000 most common alphabetic strings in the training set. Stemming was not used. We then reduced the features using Weka's implementation of correlation-based feature subset selection (Hall & Smith 1999). After evaluating several classifiers, we used Multinomial Naïve Bayes (McCallum & Nigam 1998).

	10-fold cross validation	1000 random, hold-out posts
Accuracy	0.965	0.985
κ	0.902	0.902
Specificity	0.997	0.995
Sensitivity (Recall)	0.869	0.874

Table VI-2 Classifier performance measures.

Table VI-2 presents two sets of performance measures for this classifier. The first is from a 10-fold cross validation on the full set of posts classified by human coders. Because the full set over-sampled political posts, the second evaluation classified the 1000 posts drawn uniformly at random, trained on the remaining 5,691 human-coded posts. The kappa for both tests, 0.902, was well above the benchmark that Landis and Koch (1977) propose for "almost perfect" agreement, 0.81. The sensitivity rate means that, on our random sample of 1000 blogs, 87.4% of political posts were correctly classified as political. The specificity rate means that 99.5% of posts classified as nonpolitical were in fact nonpolitical. Using the full human-coded data for training, the classifier identified 217,727 political posts and 2,136,551 non-political posts, about 10.2% political.

Though simply tallying the posts on any given blog that were classified as political and dividing by total posts gives an estimate of prevalence (percent of posts that are political) for that blog, that estimate would be biased. To see this, consider a blog that posts about politics only one percent of the time, with 10 political posts and 990 non-political ones. In expectation, 8.74 of the 10 political articles will be classified as such. Of the non-political posts, 0.5% will be incorrectly classified as political, about 4.95 posts.

Thus, when the true prevalence is 1%, the estimation procedure of simply counting the number of items classified as political, will, in expectation, yield a slightly higher true prevalence estimate, 13.95/1000 or 1.40%. The problem is even worse at the other extreme. If a blog had 100% political posts, on average, only 87.4% would be classified as political and the expected value of the estimate would be 87.4%, much lower than the true 100%. The problem with the naïve estimator is that it does not take into account the error rate of the classifier. The challenge of estimating true prevalence from observed prevalence (p) when there is a known error rate in the observation technique (the classifier in our case) has been addressed in medical statistics work (Zhou et al 2002). We generate corrected prevalence estimates (p^*) on a per blog basis according to the following:

$$p^* = \frac{p - (1 - specificity)}{sensitivity - (1 - specificity)}$$

For example, when the observed rate p was 1%, we estimate a lower adjusted rate p^* of 0.6%, and when the observed p is less than 0.5%, the adjusted p^* will be negative. When the observed rate p was 10%, we estimate an adjusted rate p^* of 10.9%. And when the observed rate p was 70%, we estimate an adjusted rate p^* of 80.0%. Because our selection of human-coded posts only included a few posts from some of less prominent blog genres, we were not able to report or use specificity and sensitivity on a per-genre basis.

We then use our corrected estimates of p^* to generate revised estimates of the total number of political and nonpolitical posts on each blog. We sum up those estimates to estimate the cumulative prevalence of political posts in various collections of blogs, including the collection of blogs that rarely if ever contain political posts. Note that when p^* is negative for a blog, that blog's contribution to the estimated cumulative total will be negative. This procedure is equivalent to adding up the number of posts for a whole collection that were classified as political or non-political to generate a value p for the

collection, and then applying the correction to generate a value $p^\star$ for the whole collection.

Classification of blogs

We then classified each of these blogs according to genre. To do this, we used Amazon's Mechanical Turk micro-task market. Each blog was shown to five users of Mechanical Turk, who were each asked to enter the title of the blog (as a check to ensure the rater looked at the blog), and to identify the category to which it belonged from a list of eight categories (with brief descriptions; Table VI-3) or to mark it as a spam blog or a blog not written entirely in English. In some cases, particularly bad workers were identified and their completed tasks were reassigned to other workers, so these tasks received more than five ratings. If the blog was still active, the workers were shown the blog in an iframe; if not, the workers were shown an iframe with text-only copies of the posts and titles. We then used the get-another-label tool[13], developed by Panos Ipeirotis, to reconcile the ratings from different Mechanical Turk workers (Sheng et al 2008). On 315 blogs labeled by both a researcher and the Mechanical Turk process, the overall kappa was 0.72. This classification process also identified 96 additional blogs that were written only partially in English or that were spam; these blogs were removed from our sample. This left 2,354,278 posts from 8,765 blogs.

---

[13] http://code.google.com/p/get-another-label/

Category	Description	n	%	Posts / week / blog Mean (stdev)	Comments / post Mean (stdev)
Diary	Individual, group, or family blog with news about life. Includes blogs that are collections of links, images, or videos that interested the author if the collection does not fit into one of the other categories.	5307	60.5%	3.62 (5.17)	2.33 (6.10)
Hobby & Fan	Blog about a particular hobby, interest, or activity (such as crafts, photography, programming, or cooking). Also includes blogs by enthusiasts of a particular TV show, celebrity, actor, movie, musical group, or sports team. Incudes travel and exercise diaries (e.g. someone who writes about running or cycling as a hobby).	2148	24.5%	5.81 (5.81)	2.20 (8.30)
Professional & sales	Blog for a trade, educational, or professional association, or containing news, tips, or advice for people in a particular career or line of work, or an official blog to promote a product, service, or event, to interact with customers, or to provide news about a business or other organization.	519	5.9%	40.1 (567.42)	2.20 (8.81)
Politics	Blog with commentary or news on issues or controversies in politics and government	422	4.8%	11.89 (18.47)	3.11 (12.88)
Religion	Blog by/about religious organizations, daily devotionals, or meditations. Does not include life diaries by people for whom religion is a big part of their life.	200	2.3%	4.06 (4.49)	2.06 (5.75)
Civic & issue	Blog that promotes a *particular* social or political change, such as an environmental organization	81	0.9%	7.85 (13.12)	1.35 (3.71)
Health & Wellness	Blog with tips, suggestions, support, or advice for health and/or wellness. Includes patient diaries and blogs with advice about exercise for health.	66	0.8%	3.75 (3.87)	2.16 (5.91)
Ethnic / cultural	Blog about a particular culture or heritage.	22	0.3%	2.27 (5.54)	1.35 (2.92)

Table VI-3 Blogs by category. Descriptions are those provided to Mechanical Turk workers.

Results

The majority of these 8,675 blogs were diary (60.5%) or hobby and fan blogs (24.5%), with political blogs accounting for less than 5% of our sample (Table VI-3). The political blogs had more posts per week than any other category except for professional and sales blogs, and also had, on average, more comments per post than other blog categories.

Not surprisingly, many blogs in our data set contained no political posts (Figure VI-1); we estimate that 30% of blogs contained no political posts and another 60% post about politics between 0% and 20% of the time, with an estimated overall prevalence (p^*) 3.0%, across blogs that each post about politics less than 20% of the time. Cumulatively, however, because there are so many of these blogs, they account for 25% of total political blog posts (Figure VI-2).
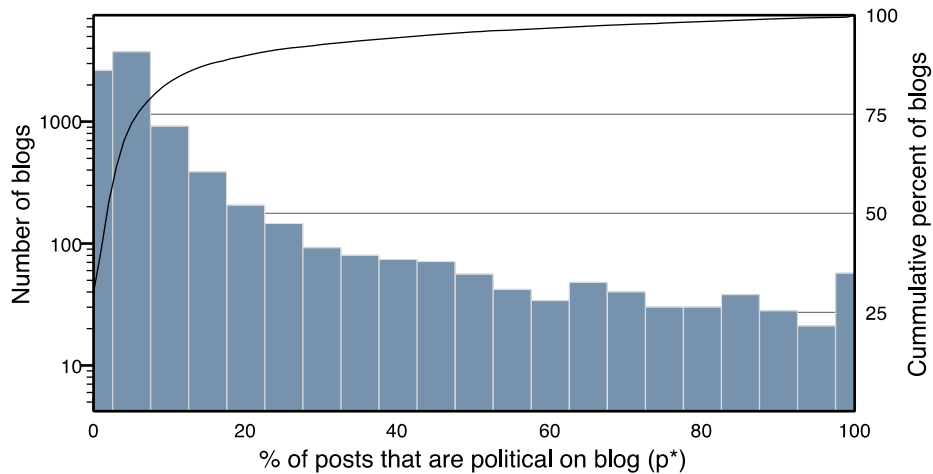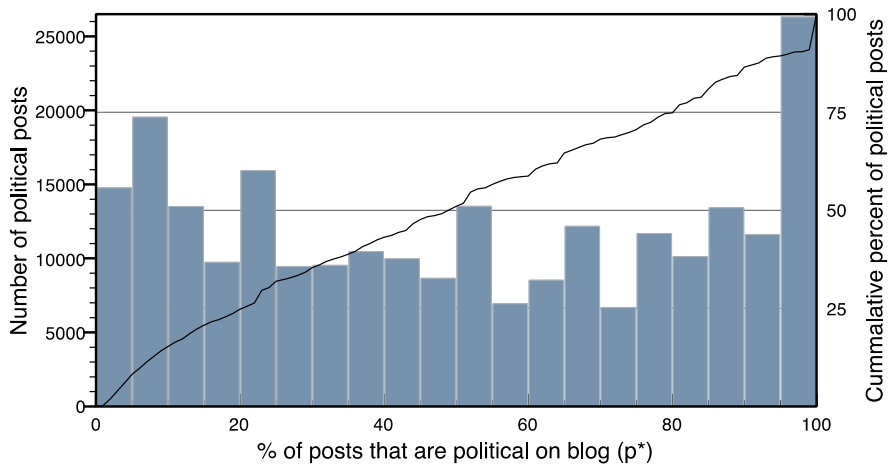
The frequency of political posts on non-political blogs fluctuated mildly, tracking national political events in the United States (Figure VI-3), most noticeably the November 2006 election and an increase leading up to and early in the 2008 primary season. The data for early 2006 are noisy because there are fewer blogs in the sample from that time. It is also possible that a portion of the measured increase in political discussion during the primary season results from the classifier having an easier time detecting political posts containing candidates' names.
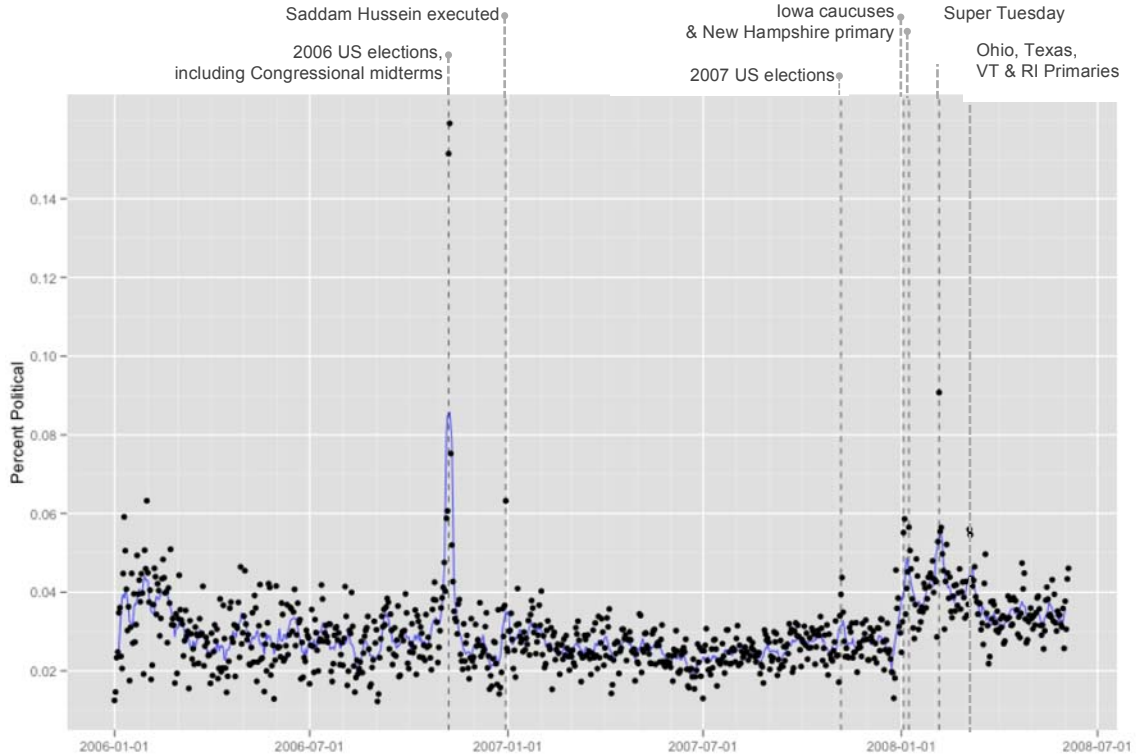
Figure VI-3 Portion of total posts, by day, that are political from blogs that are political <20% of the time. Data points per day, with seven-day two-sided moving average (blue line).

The prevalence of political posts varied across the categories (Figure 4). As one would expect, political blogs had the greatest prevalence of political posts (but even some political bloggers showed quite a bit of difficulty staying on their professed topic), followed by civic and issue blogs, which included town news sites or advocacy sites for a particular issue (e.g., environmental concerns) or a particular group (e.g., LGBT blogs). Some categories showed several more political outliers, such as diary blogs by people who had a lot to say about politics, and hobby or professional blogs where some hobbies or professions might be very connected to political issues (e.g., guns and gun control laws) while others are quite removed (e.g., scrapbooking). In general, all categories except politics and civic & issue had at least three quarters of the blogs with less than 20% political posts. On the other hand, in all categories, even personal diaries, hobby, and health & wellness, at least half the blogs had at least one political post.
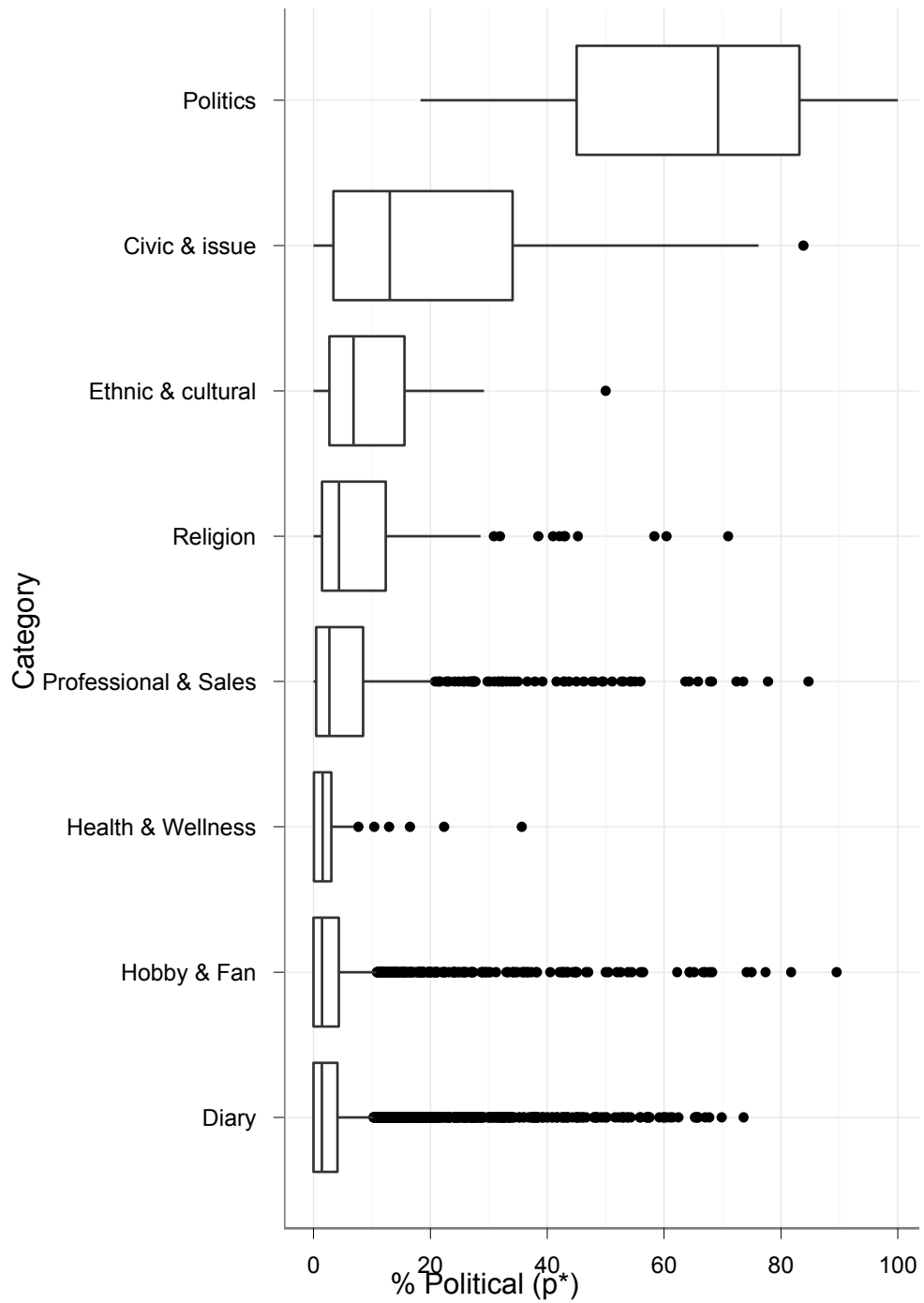
Figure VI-4 Prevalence of political posts by blog category. (box is first quartile, median, upper quartile; whiskers extend from smallest to maximum value, up to 1.5 times the interquartile range; individual blogs are shown as outliers.)

Are political posts treated as taboo on non-political blogs?

The next question we addressed using the Blogger.com sample is whether politics is treated as a taboo subject on the non-political blogs. In these non-political contexts, do blog readers simply ignore and choose not to respond to political posts? Do they discourage the author from posting such material? Though exposure to diverse points of view may be achieved simply by having diverse readers encounter political opinions on others' blogs, interaction is required for a deliberative conversation to occur.

Many of the authors seem to be self-conscious about their decision to bring politics into the non-political spaces, beginning or ending the post with statements such as "Rambling, Uninformed, Brutish Political Rant of the Week," "forgive my outrage," "okay, time to get off the soapbox," or "please excuse my rant." This appeared to be particularly true for posts that had politics as their main topic. Future work should examine, in detail, the prevalence and nature of such introductions or warnings before making a political post in a non-political blog.

Despite this written hesitation from the authors, political posts on nonpolitical blogs receive as many comments as nonpolitical posts on the same blog. These posts also did not appear to be getting more responses just because they are an exception to the blog's normal content: political posts on political blogs also receive more comments than the non-political posts, suggesting that the political posts on the nonpolitical blogs do not receive more comments simply because they are about a different topic than other posts on the blog.

Table VI-4 shows the means for each group, but this analysis is problematic because it treats posts from the same blog as independent, when they are, in fact, not. To better model the comment count, we computed a negative binomial regression model for a post's expected comment count based on whether the post was political or not (an indicator variable), whether the blog was categorized as political or not (an indicator

variable), and the interaction effect between whether the post was political and whether the blog was political. Because this analysis compared the comment counts for the different posts on the same blog, and because one of the explanatory variables (whether a blog is political) was constant for each blog, we used random effects methods on this panel data set (Table VI-5). The regression model indicates that political posts on non-political blogs receive about the same number of comments as non-political posts on those blogs.

| | | Mean comments on | |
	N	Political Posts	Nonpolitical posts
Non-political blogs	8343	2.506	2.250
Political blogs	422	3.489	2.658

Table VI-4 Mean comments per post by post and blog type. Note that the means presented in this table are highly influenced by more prolific blogs, while blogs with few posts have little influence.

	IRR	Std err	*p*-value
Political post	1.006	0.004	ns
Political blog	0.273	0.018	<0.001
Political post * political blog	1.297	0.011	<0.001
Log-likelihood		-3480872.9	

Table VI-5 Negative binomial regression model for the expected comments on a post, given whether the post was political and whether the blog was political. n=2,354,278 posts across 8,765 blogs. Wald χ^2 = 44055.34, $p < 0.001$.

According to the regression model, the political blogs in our sample tended to receive fewer comments overall, but received more comments on their political posts than on their non-political posts. This appears to be contrary to the means presented in Table VI-4. One reason for this is that the means in Table VI-4 are strongly influenced by the blogs with the most posts.

In addition, looking at the mean comments per blog, comparing the distributions for political and non-political blogs, it turns out that, for most parts of the distribution, political blogs get fewer comments per blog. In the 95th to 99th percentiles of the respective distributions, however, the political blogs get about twice as many comments per post as

non-political blogs. This increases the means, but has little effect on the maximum likelihood estimation of the coefficient for political blogs, because only a small percentage of political blogs have higher comments per entry.[14]

We were also curious whether political posts increased the expected number of authors in each comment thread, or whether the number of authors remained the same or even decreased, with participants engaging in more back-and-forth discussion. We repeat the previous analysis with the author count (number of authors in the comment thread) as the dependent variable, adding comment count as an independent variable, and using only posts with comments (Table VI-6). For the purpose of this analysis, all anonymous comments on a thread were treated as written by the same author. For posts with the same number of comments, political posts appear to elicit comments from slightly more authors (4.5% more), regardless of the type of blog.

	IRR	Std Error	p-value
Political post	1.045	0.00346	<0.001
Political blog	0.981	0.01316	ns
Political post * political blog	0.998	0.00644	ns
Comment count	1.009	0.00001	<0.001
Log-likelihood		-2124540.5	

Table VI-6 Negative binomial regression model for the expected number of authors appearing in a post's comments, given whether the post was political, whether the blog was political, and the number of comments. n=1,050,515 posts with comments across 8,765 blogs. Wald $\chi^2 = 294214.08$, $p < 0.001$.

Finally, we tested whether anonymous comments were more prevalent on political posts on non-political blogs, something that might happen if people wanted to speak up but did not want to reveal their identity (potentially putting their relationship at risk). For

---

[14] It was, however, sensitive to one outlying news blog, classified as political, that had more than 25,000 posts and more than six comments per post, but relatively few posts (25%) classified as political. Removal of this blog resulted in IRR estimates of 0.47 for a political blog (p<0.001) and 1.08 for a political post on a political blog (p<0.001). These estimates are more consistent with estimates from models based on 20% sub-samples of blogs drawn from our sample.

The estimates of the effect of political posts in non-political blogs were quite stable in all models, consistently indicating that political posts in non-political blogs get very slightly more comments than non-political posts in those blogs.

this analysis, we repeated the above procedure, but restricted it to posts, with comments, on nonpolitical blogs, and added the author count (rather than comment count) as a control (Table VI-7). We see a 15% increase in the expected amount of anonymous comments for political posts on nonpolitical blogs.

	IRR	Std Error	p-value
Political post	1.152	0.0098	<0.001
Author count	1.013	0.0001	<0.001
Log-likelihood		-625029.01	

Table VI-7 Negative binomial regression model for the expected number of anonymous comments appearing in a post's comments, given whether the post was political, whether the blog was political, and the number of authors. n=974,528 posts with comments across 8,343 non-political blogs. Wald $\chi^2 = 20830.1$, $p < 0.001$.

CONTENT ANALYSIS

Because many of the political posts on non-political blogs also discussed other subjects, we sought to measure whether the comments on nonpolitical blogs engaged the political content of the post, talked about something else, or discouraged the post's author from discussing politics on their blog. Unfortunately, the classifier for whether a post was political or not was impractical to use on the comments: the comments tended to be briefer, and, taken out of context, were likely to produce unreliable results. Instead, we sampled 250 posts and associated comment threads at random from the political posts on non-political blogs that had at least one comment. Researchers first verified that each of these posts had some political content, and coded it for whether it was about other topics as well. For those posts that were indeed political, we coded each responding comment for:

- whether it was spam,
- whether it engaged the political content of the post,
- whether the commenter agreed, disagreed, or neither with the blogger's political position (according to the criteria for agreement in Gilbert et al 2009), and
- whether the author said the blog post's content did not belong.

In total, we coded 1188 comments on 244 political posts (the other 6 had been misidentified as political by the automated classifier); 23 of these comments were removed as spam comments. 56 comments on 42 posts were coded by at least two researchers; inter-rater reliability on each of the coding categories is shown in Table VI-8.

Decision		% agreement	κ
Post	Had some political content?	100%	1.00
	Had other, non-political topics?	100%	1.00
Comment	Spam?	98%	0.87
	Engaged political content of post?	93%	0.79
	Agreed, disagreed, or neither with post's author?	91%	0.84
	Said the blog post's political content did not belong?	100%	1.00

Table VI-8 Inter-rater reliability on 56 comments on 42 blog posts.

Of these 244 political posts on nonpolitical blogs, 60 (25%) also talked about at least one other subject that was not politics. The posts had an average of 4.8 non-spam comments per post, which drops to 990 comments (an average of 4.1 per post) once comments by the blog post author are excluded (Table VI-9).

		n	%
Engaged political content of post		823	83%
Among comments engaging with political content			
Expressed	agreement	334	41%
	neither / balanced	370	45%
	disagreement	119	14%

Table VI-9 Analysis of 990 comments by people other than the post author and not by spammers on 244 political posts on nonpolitical blogs.

Because many of these posts contained other topics, and because even a political post might prompt a friend to post a comment about simply catching up, we assessed how many of these comments engaged the political content of the original blog post or a subsequent comment. 961 (83%) of the 1165 total comments, and 823 (83%) of the comments by people other than the blog post author, engaged the political content in some way. While a more thorough analysis of the discourse quality is beyond the scope of

this paper, we note that these comments included a mix of thoughtful critiques of policies, personal stories, advocacy for candidates and issues, and ad-hominem attacks on or crass jokes about the appearance or behavior of politicians and their family members. Only two comments (0.2%) suggested that politics did not belong on the blog.

We next looked at the rates of agreement and disagreement among the political comments on these nonpolitical posts, considering just the comments by someone other than the blog author. Of these comments, 334 expressed agreement with the post's author (41%), 119 expressed disagreement (14%), and 370 (45%) expressed neither disagreement nor agreement or were balanced between the two. This deviates only slightly from the agreement to disagreement ratio that Gilbert et al found among political blogs (47% agreement, 13% disagreement, and 40% neither), though it does so in the direction of slightly more disagreeing or neutral comments (an agreement to disagreement ratio of 2.9 rather than 3.6).

Discussion

The volume of political discussion on non-political blogs from Blogger.com is substantial, both in the posts and the comments on these posts. This offers some support for Wojcieszak and Mutz's finding that people report more exposure to cross-cutting political opinions in non-political online spaces than in political ones (2009), in that political discussion does exist on non-political blogs. This work adds an important qualifier, though: even if people are encountering challenging or disagreeable political opinions when reading non-political blogs, they are at best only slightly more likely to voice that disagreement in the blog comments than commenters on political blogs. A variety of factors may contribute to the high agreement to disagreement ratio observed in the comments, including political homophily among social networks (Goel et al 2010) and a tendency of those who disagree to stay quiet in order to maintain social harmony (Noelle-Neumann 1993).

We might expect similar volumes of political discussion in at least some other online non-political spaces – other blogging services or social network sites such as Facebook and Twitter. A Pew Internet and American Life Project survey (2011) shows that a portion of adults are publicly posting political content and supporting or following political figures and issues on Twitter and other non-political social network sites, and even more are learning about their friends' political preferences as a result of these actions. One important direction for future research is to investigate how different formats and design features differ in their affordances for political discussion. For example, it may be that blogs are treated as a zone of personal expression and so it is socially acceptable for the author to write occasionally about politics, while the same freedom may not occur in forums or email lists that have formed around non-political topics.

If many online non-political spaces serendipitously expose their participants to diverse political information, they offer a counter to the Internet's increased affordances for crafting an agreeable political information diet. Yes, people may choose news websites or configure news aggregators to show opinions with which they agree, but they will still be participating in other online spaces that bring together people with diverse views and who sometimes talk about those views. The demise of the inadvertent audience (Prior 2007, Bennett & Iyengar 2008) has likely been overstated. Though no longer coupled with broadcast media, people are serendipitously and inadvertently being exposed to political news and information across the many online spaces in which people interact and share information. So long as designers of social sites do not begin naïvely trying to filter out disagreeable political posts or opinions in personalized feeds (a design choice that one might fear from a reading of Pariser 2011), these non-political spaces may continue to keep selective exposure preferences in check.

While we have shown that at least some non-political spaces contribute to online political discourse, and that they do include at least some diversity in views, we have yet to evaluate whether these non-political spaces yield more deliberative political discussions

than political spaces. I believe that better understanding and characterizing the political

discussion occurring in online, non-political spaces is a critical area for future research.

CHAPTER VII

Summary

In my dissertation research, I set out to (1) better understand preferences for diversity in the political information people access online and to (2) develop and assess techniques to select and present challenging items in ways that are appealing to challenge averse readers. I also explored whether non-political spaces contribute to political discourse online, if such spaces might be important sources of exposure to counter-attitudinal political viewpoints, and whether people choose to engage such discussion when they encounter it.

      I worked on the first two goals in the context of online news aggregators such as *Digg* and *Reddit.* To my surprise, I found that some people show diversity-seeking behavior and others show challenge-averse behavior – there is no average human who can be described by one theory or the other (Chapter IV). Some simple presentation techniques – highlighting agreeable items or highlighting the agreeable items and placing them first in the news aggregator's results – did not increase subjects' satisfaction with lists containing challenging items. In Chapter V, I sketch a design space for techniques to present feedback about the political leaning of the items one reads, including foresight, "currentsight", and hindsight widgets, and a design for a study to test at least a hindsight widget. To complement work on presentation techniques, I developed the Sidelines algorithm (Chapter II), designed to select more representative result sets from user votes, and evaluated it against the approval voting algorithm on existing data sets and in an

online experiment, and found that the Sidelines algorithm does increase the diversity of result sets.

To address the final goal, I studied an example of non-political online spaces, where diverse discussion may already occur (Chapter VI). Analysis of a random sample of blogs from *Blogger* revealed that, at least on that site, half of the total political posts occur on blogs that post about politics less than half of the time. Moreover, when readers encounter these posts, they do not treat them as taboo, and they engage with the posts' political content. These results support recent survey work by Wojcieszak and Mutz (2009) that found that people report they are most likely to encounter cross-cutting political opinions on non-political websites, not on political or news sites. These inadvertent or serendipitous encounters with political viewpoints other than one's own are likely to continue to limit the effects of selective exposure preferences, even if people do increasingly access niche or personalized news websites when they seek out political news. Further work should explore the role of these non-political spaces in online political deliberation and attitude formation in much more detail. In particular, researchers should code threads in non-political and political spaces for characteristics such as signs of listening and openness to changing opinions, the number of arguments represented, the number of positions represented, and tone or civility.

This research advances our knowledge of how the preferences predicted by selective exposure theory translate to behavior online, a topic important to political theorists and designers and moderators of online political news and opinion sites and non-political spaces, and emphasizes that selective exposure research must consider individual differences. It has also resulted in at least one tool (the Sidelines algorithm), which could be used in practice today.

CHAPTER VIII

What's next?

In the introduction, I described five research challenges for understanding, and perhaps decreasing, the challenge-averse behavior that selective exposure theory predicts for how people access political news and information online. Four challenges address political news aggregators: (1) understanding preferences and behaviors for accessing political information online, (2) exploring ways to nudge challenge-averse individuals to read or consider more diverse political news, (3) studying techniques to help people engage with diverse points of views, and (4) evaluating the actual outcomes of achieving these diversity goals. The fifth challenge involves understanding the political information sharing and discussion that happens in *non*-political spaces online.

My studies touch only three of these vital research areas. In this chapter, I review contemporary research that complements my work and recent trends in access to news and political information. I then suggest future directions and potential next steps for the study of selective exposure preferences and behavior on the Internet.

Contemporary research

Others have made concurrent progress on understanding existing preferences and behaviors (e.g., Wojcieszak & Mutz 2009, Park 2011) and on the selection and presentation challenges (e.g., Park 2009, Oh 2009). Some of this work confirms or complements my results, particularly with respect to individuals' preferences for exposure

to diverse points of view and the online spaces in which that exposure may occur. The research on techniques for selecting and presenting diverse political items has identified some approaches that appear to be more promising than the ones I have designed and evaluated to date.

The *NewsCube* project, from researchers at KAIST (Park 2009), is particularly worth noting. *NewsCube* groups news articles on a topic into different clusters ("aspects"). The system uses an unsupervised approach to build clusters: it first extracts keywords in each article in a topic, and then uses hierarchical agglomerative clustering to build article clusters based on keyword similarity. Readers select and read an article on a topic, but are also shown that there are other clusters – and thus likely other perspectives or viewpoints – available to explore on the topic. When presented with this interface, readers explored more topics than they did using the *Google News* interface, in which articles on any given topic are not further clustered. Showing these clusters seems to have increased subjects' curiosity about the different aspects to a story, and it appears that this technique can, at least in the short term, decrease the effects of selective exposure preferences.

Other researchers have tackled the third broad area: encouraging people to engage with and consider the diverse opinions they do encounter online. Kriplean et al developed two systems to promote "listening" online. *Reflect* adds an additional layer of discussion to comment systems in tools such as the blogging platform *WordPress* (Kriplean et al 2012a). Rather than being general purpose, though, this level of comments is meant for summarizing others' posts, and the original commenter can mark the quality of a summary up or down. The authors hoped that, by giving people a clear space in which to "listen" and a way for commenters to acknowledge good listening, people would focus less on expressing their own opinion and more on considering others' opinions and then engaging them in discussion. In two field deployments – one on the tech news site *Slashdot* and the other on *Wikipedia,* the authors observed at least some success. In the

second system, *ConsiderIt,* the researchers developed an online voters' guide in which users could craft, individually and collaboratively, pro and con lists for various ballot initiatives (Kriplean et al 2012b). In a field deployment, people expressed positions that included both pros and cons, and they engaged with positions different than their own.

In a very different approach, Sukumaran et al (2011) conducted two lab experiments to test whether it is possible to foster more thoughtful commenting and participation in online discussion forums by priming thoughtful norms. The first experiment tested the effects of the behavior of other participants in the forum (social proof and/or modeling). The dependent variables were comment length, time taken to write the comments, and number of issue-relevant thoughts. Not surprisingly, being exposed to other thoughtful comments led people to make more thoughtful comments themselves. The second study tested effects of visual, textual, and interaction design features on the same dependent variables. The manipulations included a more subdued vs. more playful visual design, differing CAPTCHAs (words positively correlated with thoughtfulness in the thoughtful condition and words negatively correlated with thoughtfulness in the unthoughtful condition), and different labels for the comment box. The design intended to provoke thoughtfulness led to more thoughtful comments, suggesting that it is possible, at least in the lab, to design sites to prompt more thoughtful discussion.

Increasing role of the Internet for political news & information

Online spaces and their roles in daily life have also evolved since I started this research. Overall, they have become increasingly prominent. In 2010, 61% of Americans said they use the Internet to access news – more than radio and newspapers, and just behind television – and more than a quarter of Americans used smartphones to access news on the go (Purcell et al 2010). By 2012, online sources were the primary news source for twice as many Americans as were print newspapers (Pew 2012). 70% of Americans, however,

find the current volume and variety of news and sources overwhelming, and so, despite a range of choices, the majority of Internet news users access no more than two to five news sites (Purcell et al 2010). Curiously, the news audience actually seems to be shrinking somewhat – in 1996, 14% of Americans reported accessing no news in a typical day, while this had increased to 17% by 2010.

While blogs were once the hotbed of discussion and conversation online, much more time and attention is now spent in spaces that look quite different. In contrast to blogs' affordances for a single author or group of authors setting an agenda or conversation and then having commenters react and discuss, social sites such as Facebook and Twitter aggregate posts from many users in a single feed, and replies have prominence similar to the original post. Though these sites currently account for a small portion of how Americans are accessing news content – in 2012, only 9% of digital news consumers say that they very often follow news recommendations from Facebook or Twitter – this number is increasing (Pew 2012). Additionally, other, important information is transmitted through these channels. 37% of social network site users report posting political content at least occasionally, and 38% have discovered, by reading posts to social network sites, that friends' political leanings were different than they believed (Rainie & Smith 2012). Recently introduced features will likely increase the importance of this channel for discovery of news, opinions, and others' political preferences and reading choices. One such feature, Facebook's *Social Reader,* automatically posts, in one's feed and the new news ticker view, links to stories friends have recently read. Twitter now embeds story previews for articles that people have shared in the feed, on both the website and their desktop and mobile applications as part of its "expanded tweets" feature.

A future research agenda

The sum of my work still does not fully describe individuals' preferences and behaviors for online news access, does not completely characterize online political discussion in the

many spaces in which it occurs, and does not prescribe a clear set of guidelines for designing spaces that expose people to more diverse political content. It does, however, advance the understanding of preferences and behaviors and help guide future design, engineering, and research directions. In this section, I will review what I believe are some important next steps.

Preferences

Selective exposure theory argues that people are challenge-averse and will avoid sources that are likely to present them with counter-attitudinal information. The study in Chapter IV shows that there are individual differences in these preferences. Yes, many subjects appeared to be challenge-averse, but a number appear to be diversity seeking. Knowing that there are individual differences opens the door to additional questions, including more nuanced understandings of preferences and how they may vary across time, topics, or context.

For an individual's instantaneous preference, does the tone or rhetorical style of a counter-attitudinal item affect their receptiveness to it? It will also be important to know whether challenge aversion or diversity seeking is an inherent trait of an individual, or whether this preference varies and, if so, according to what. An individual's preferences and behavior may vary from topic to topic, how central an issue or position is to their identity, or whether or not their party or interest group currently has power. Future work should consider such factors as well.

News aggregators

The studies described in Chapters II-V focused on changes to news aggregators that present and select content based on reader submissions and votes. I am skeptical of the continued value of studying and designing primarily for these use cases, based on recent Pew data about where Americans access online news: only 7% of Americans access a news posting, ranking, or rating site such as *Digg* or *Reddit* on a daily basis, while 68% visited

aggregators such as *Google News* or *AOL News* (Purcell et al 2010). I caution against developing solutions that work only on user-curated sites such as *Digg* and *Reddit,* as such a trajectory may solve one problem only to find that the princess is, unfortunately, in another castle. Instead, researchers should focus on the large opportunity to have impact by developing solutions that will also work on sites such as *Google News* or across the web.

Many of the techniques colleagues and I considered for *Digg* and *Reddit*, particularly the presentation techniques, can be applied to sites such as *Google News* with no or small changes. If clicks or likes are used as a proxy for user votes, Sidelines and its ilk could be used in selecting items on a broader range of sites. Next, I describe some approaches for selection, presentation, or some combination that merit future study.

Beyond algorithms that rely on user votes (or proxies for votes, such as links, clicks, "likes", or comments and comment sentiment), classification and clustering methods based on article text and features such as source and author hold potential. I had initially been skeptical of such approaches. The topics receiving political attention, political figures, and even different parties' positions on issues can change over time, and thus a steady stream of hand-coded training data would be needed to correctly classify breaking news. Two sets of research have changed my opinion.

First, article clustering need not match pre-defined categories to be useful. *NewsCube* shows that unsupervised clustering can group articles on emerging topics, and that this clustering can add value– even if the clusters are not labeled according to, or aligned with, traditional party positions, or labeled at all. Indeed, it is possible that presenting these clusters as other perspectives on the topic, rather than as corresponding to a partisan viewpoint, was actually an asset in encouraging exploration of more aspects to a story, as signals such as "left" or "right" labels might serve only to heighten readers' aversion to and avoidance of counter-attitudinal clusters. At the University of Michigan, Sidharth Chhabra is currently working on replicating the *NewsCube* work in a way that is

suitable for a long-term field trial, and it could be available as a news aggregator on which researchers can conduct field experiments. Second, as discussed in Chapter V, researchers have shown that it is possible to use article text and other features, such as URL, to reliably classify articles as political or not, and, if political, as liberal or conservative (e.g., Oh et al 2009, Gentzkow & Shapiro 2010, and Dehghani et al 2011).

A combination of these techniques should make it possible to build a news aggregator on which one can test several different selection and presentation techniques. Such a news aggregator would also enable researchers to study whether clustering continues to affect reading behavior over time, or whether there is simply a novelty effect, and also to conduct experiments on the effects of features such as visualizations showing the percentage of topics explored, more normative information such as how an individual's readership compares to other readers', or progress against a goal a reader or the designer sets (such as balancing the character in the hindsight widget presented in Chapter V). One could imagine a large-scale effort to test, in the field, several selection and presentation techniques, as well as other features designed to influence reading behavior.

Selection: Clustering vs. Partisan (e.g., liberal-conservative). Does unsupervised (and thus unlabeled) clustering encourage more exploration of diverse viewpoints than sets of articles labeled as liberal, conservative, independent, and so on? What if both sets were unlabeled, i.e., how much do the actual clusters matter versus their labels? Also, while my studies selected and displayed articles according to the classic liberal-conservative spectrum (as have studies such as Oh 2009), this convenient simplification only captures one of the many dimension on which articles may vary. Other dimensions that selection techniques might (and arguably, should) eventually consider include whether an author argues from a fact-based or value-based perspective, how much an article differs from the reader's position, whether arguments are supported with statistics

or stories, and the general tone and affect of an item. All of these, and more, may substantially affect how receptive an individual is to articles.

I also remain curious about social selection techniques: selecting stories popular among people who are ideologically different, but perhaps socially connected to the reader, or similar on certain key demographics, and then revealing those connections or similarities when displaying the article. I will discuss some of the challenges with social selection of articles later in this chapter.

Visualization techniques. In Chapter V, I outlined a range of visualization techniques that should be assessed. This includes techniques that reveal the predicted bias of an item before a user chooses to visit it (foresight) or as a user reads it ("currentsight") and techniques that present the overall balance of their recent reading history (hindsight). As illustrated with the BALANCE man, these visualizations could communicate what the reader *should* do, not only what they are doing or have done. Injunctive norms can be added to foresight or hindsight widgets, such as by overlaying a negative emoticon (☹) to a currentsight widget when an imbalanced reader is reading yet another affirming item, or a positive emoticon (☺) when a reader chooses a counter-attitudinal item.

Commitment mechanisms could also be explored for increasing the diversity of news that people read. A variety of commitment mechanisms may help people live up to their "better selves" in their reading behavior. What if a news aggregator asked people to be more balanced in their reading, or to explore more topics than last week? What if a reader agreed to read some articles recommended by someone ideologically different in exchange for their counterpart reading articles they recommend? A great deal of research on commitments and targets could inform the design of these features to make them, potentially, particularly effective – for example, by having readers state not just a diversity goal, but how they plan to achieve it or why they think it is important at the time of their commitment (Locke & Latham 2002). Commitment contracts, commonly implemented

as "I will give money to charity *x,* which supports a cause I do not believe in, if I do not meet my goal" (Schelling 1984, Ayres 2010), may be particularly apropos for this goal: either read about the other side's viewpoints or be stuck giving money to support them.

Different processing systems. Many of these future investigations may benefit from consideration of and investigation into the different cognitive systems that people use to process information (Kahneman 2011). From work in other fields, we know that different types of messages can persuade differently. Some people are more persuaded by messages that use statistics than narrative messages, though results are mixed (Allen & Preiss 1997). Statistical messages engage heuristic and systematic processing, while narrative messaging engages only heuristic processing (Kopfman et al 1998). This may apply to the political space: do particular article types and presentation techniques trigger different processing mechanisms? Does that change how amenable or receptive people are to those articles, particularly attitude-challenging ones? If so, understanding which presentation methods and which arguments appeal to different systems may help guide design choices for which political news to present and how to present it.

New spaces: Facebook and Twitter

As mentioned previously, social network sites such as Facebook and Twitter represent a small but growing channel for political news discovery and sharing, and for discovery of friends' and family members' political opinions. Could these spaces become the new coffee shops and bowling leagues, in which diverse (or at least somewhat diverse) people connected for other reasons come together and share a variety of political views and opinions?

The evidence here is mixed. Goel et al (2010) reported that though there is political homophily in Facebook networks, there is also more diversity than individuals believe. Explicit sharing of political affiliations or views in one's profile, choosing to post political stories to one's stream or feed, or "frictionlessly sharing" the stories one reads

through tools such as the Facebook *Social Reader* can bring that diversity to the surface. An important question, then, is what people do once these topics come up. On *Blogger*, I found that people do engage with political posts in non-political spaces, and even post disagreements with the author. I do not, however, have a measure for how many people choose to tune out such posts, or even stop reading a blog altogether upon discovering that the author holds views that are different from their own, and so I can only measure the positive signal: that is, how many people choose to engage by commenting on the post and how they choose to engage, and how that compares to non-political posts.

Others have studied what happens when people encounter counter-attitudinal posts from Facebook and Twitter connections. Rainie and Smith report that 66% of social network site users report always ignoring posts with which they disagree, 28% report responding, and 5% report choosing whether or not to reply depending on circumstances. More troublingly, 9% of social network site users report blocking, unfriending, or hiding someone for political posts with which they disagreed or found offensive, and 5% report blocking, hiding, or unfriending someone for political posts they worried would offend other friends. 8% of SNS users report blocking, unfriending, or hiding someone because they argued about political issues on the site with the user or someone the user knows. Though the majority of people unfriended over political differences are weak ties such as a distant friend or acquaintance, one third of people who have blocked or unfriended another user over political posts report having done so to a close friend or family member. These self-reported rates are possibly depressed by confounds such as the social desirability effect, but are consistent with other studies that identify politically polarizing posts as one of the top reported reasons for unfriending or unfollowing (Sibona & Walczak 2011, Kwak et al 2011). It is possible, then, that features that increase exposure to challenging political information in the short term can promote the destruction of ties with others who have differing views over the long term.

As a research community, we should seek to understand the effects of increased visibility of political preferences and news consumption on exposure to diverse points of views and on relationships. This will also be important to the social computing community's practitioners: when Facebook encourages features such as *Social Reader*, or Google decides to suddenly start broadcasting all of a user's *Reader* shared items to *all* of that user's contacts, by default, they do so at what cost to relationships? While some extreme views may very well warrant unfriending, it seems that most unfriending resulting from a revealed, politically challenging view, is a loss for individuals' opportunities to learn about other views and perspectives. If this sort of unfriending were to become commonplace, it threatens to reduce the amount of counter-attitudinal information to which people are exposed, and it would be unfortunate if one of the primary outcomes of tools that facilitate easier sharing would be people further shaping their networks to exclude encounters with challenging views.

We should, then, seek to find techniques that allow people to express their views without damaging relationships, and that might encourage empathy, engagement, and understanding, rather than revulsion and disconnection. Such techniques might involve selecting articles that are not *too* challenging to appear in the feeds seen by one's friends (using some of the selection dimensions discussed earlier in this chapter), or presenting more challenging articles using veiled viral marketing (Hansen & Johnson 2012) – that is, showing that a story is popular with one or more people in your network without revealing who (and thus making unfriending or unfollowing that friend, based on the veiled post, impossible).

Civic engagement, idea formation, and side effects

Finally, the fourth pillar – evaluating the actual outcomes of achieving these diversity goals – must remain on the research horizon. Because the techniques I tested in Chapter IV are of limited efficacy – or none – it remains premature to test them and their effects

on awareness and engagement in the wild. This is, however, still an important agenda item going forward. As techniques to select, present, and encourage engagement with diverse points of view mature, we should combine different sets of best practices and most promising techniques into products, deploy them for months or longer in a controlled field setting, and measure and assess outcomes. Though political theorists argue that deliberation involving diverse points of view is good for democracy, this is not the only prediction. Others find that exposure to challenging viewpoints, particularly in conversational environments, can lead people to actually become more disengaged (Scheufele et al 2004, Wojcieszak et al 2010, Dilliplane 2011). Are some settings and techniques for introducing more diverse points of view more effective at producing desirable outcomes than others? How do individual differences interact with context and features to predict how someone will act after viewing challenging, agreeable, or diverse political articles?

I hope that the studies presented in this dissertation, as well as my future research and contemporary and future work by other researchers and practitioners, will help us better understand individuals' preferences and behavior for accessing political information online, what they do when they encounter that information, and what designs can nudge people toward pro-social outcomes. Through such work, we can help the Internet achieve its potential benefits for democratic society, rather than deliver on its dystopian possibilities.

REFERENCES

Adamic, L and Glance, N. 2005. The Political Blogosphere and the 2004 US Election: Divided They Blog, *Proceedings of the 3rd international workshop on Link discovery*: 36-43.

Allen, M and Preiss, R. 1997. Comparing the persuasiveness of narrative and statistical evidence using meta-analysis, *Communication Research Reports* 14(2): 125-131.

Ayres, I. 2010. *Carrots and Sticks.* New York: Bantam.

Barron, JA. 1969. An Emerging First Amendment Right of Access to the Media, 37 *George Washington Law Review* 487.

Benhabib, S. 1996. Toward a Deliberative Model of Democratic Legitimacy, in *Democracy and Difference: Contesting the Boundaries of the Political*, 67-94. Princeton, New Jersey: Princeton University Press.

Bennett, WL and Iyengar, S. 2008. A New Era of Minimal Effects? The Changing Foundations of Political Communication, *Journal of Communication* 58: 707-731.

Berelson, B and Steiner, GA. 1964. *Human behavior: An inventory of scientific findings.* New York: Harcourt, Brace & World.

Bishop, B. 2008. *The Big Sort: why the clustering of like-minded America is tearing us apart.* New York, New York: Houghton Mifflin Company.

Brams, S and Fishburn, P. 1978. Approval Voting, *American Political Science Review* 72(3): 831-847.

Brin S and Page, L. 1998. The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems* 30(1): 107-117.

Broadbent, DE and Broadbent, MHP. 1981. Recency Effects in Visual Memory, *Quarterly Journal of Experimental Psychology* 33(A): 1-15.

Brown, R. 1986. *Social Psychology, the second edition.* New York: Free Press.

Browne, GJ and Pitts, MG. 2004. Stopping rule use during information search in design problems, *Organizational Behavior and Human Decision Processes* 95(2): 208-224.

Carbonell, J and Goldstein, J. 1998. "The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries," *Proceedings of SIGIR* 1998: 335-336.

Chamberlin, JR and Courant, PN. 1983. Representative Deliberations and Representative Decisions: Proportional Representation and the Borda Rule, *American Political Science Review* 77(3): 718-733.

Cialdini, RB. 2009. *Influence: Science and Practice* (5th Edition). Boston: Pearson.

Cialdini, RB; Kallgren, CA; Reno, RR. 1991. A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior, *Advances in Experimental Social Psychology* 24: 201-231.

Cohen, J. 1960. A coefficient of agreement for nominal scales, *Educational and Psychological Measurement* 20(1): 7–46.

Cohen, J. 1989. Deliberation and Democratic Legitimacy, In *The Good Polity: Normative Analysis of the State*, edited by A Hamlin and P Petit. New York, New York: Blackwell.

Consolvo, S; McDonald, DW; Toscos, T; Chen, M; Froehlich, JE; Harrison, B; Klasnja, P; LaMarca, A; LeGrand, L; Libby, R; Smith, I; Landay, JA. 2008. Activity Sensing in the Wild: A Field Trial of UbiFit Garden, *Proceedings of CHI* 2008.

Dehghani, M; Gratch, J; Sachdeva, S; Sagae, K. 2011. Analyzing Conservative and Liberal Blogs Related to the Construction of the Ground Zero Mosque, *Proceedings of CogSci 2011:* 1853-1858.

Dewey, J. 1927. *The Public and Its Problems.* Athens, OH: Swallow Press.

Diakopoulos, NA and Shamma, DA. 2010. Characterizing Debate Performance via Aggregated Twitter Sentiment, *Proceedings of CHI* 2010.

Dilliplane, S. 2011. All the News You Want to Hear: The Impact of Partisan News Exposure on Political Participation, *Public Opinion Quarterly* 75(2): 287-316.

Donnellan, MB; Oswald, FL; Baird, BM; Lucas, RE. 2006. The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment* 18: 192-203.

Esterling, KM; Fung, A; Lee, Taeku. n.d.. How Much Disagreement is Good for Democratic Deliberation? The CaliforniaSpeaks Health Care Reform Experiment. Working paper as of 10 January 2012.

Eliasoph, N. 1998. *Avoiding Politics: How Americans Produce Apathy in Everyday Life*. Cambridge University Press.

Festinger, L. 1957. *A theory of cognitive dissonance*. Evanston, Illinois: Row & Peterson.

Fogg, BJ. 2002. *Persuasive Technology: Using Computers to Change What We Think and Do*, San Francisco, California: Morgan Kaufmann.

Freedman, JL. 1965. Confidence, utility, and selective exposure, *Journal of Experimental Social Psychology* 2: 778-780.

Frey, D. 1986. Recent Research on Selective Exposure to Information, *Advances in Experimental Social Psychology* 19: 41-80.

Froehlich, JE; Dillahunt, T; Klasnja, P; Mankoff, J; Consolvo, S; Harrison, B; Landay, JA. UbiGreen: Investigating a Mobile Tool for Tracking and Supporting Green Transportation Habits, *Proceedings of CHI* 2009.

Garrett, RK. 2009. Echo chambers online?: Politically motivated selective exposure among Internet news users, *Journal of Computer Mediated Communication* 14(2): 265-285.

Garrett, RK and Resnick, P. 2011. Resisting Political Fragmentation on the Internet, *Daedalus* 140(4): 108-120.

Gentzkow, M and Shapiro, JM. 2008. Competition and Truth in the Market for News, *Journal of Economic Perspectives* 22: 133-154.

Gentzkow, M and Shapiro, JM. 2010. What Drives Media Slant? Evidence from US Daily Newspapers, *Econometrica* 78(1): 35-71.

Gentzkow, M and Shapiro, JM. 2011. Ideological Separation Online and Offline, *Quarterly Journal of Economics* 126: 1799-1839.

Gilbert, E; Bergstrom, T; and Karahalios, K. 2009. Blogs Are Echo Chambers: Blogs Are Echo Chambers, *Proceedings of HICSS* 2009.

Goel, S; Mason, W; Watts, DJ. 2010. Real and perceived attitude agreement in networks, *Journal of Personality and Social Psychology* 99(4): 611-621.

Golbeck, J and Hansen, DH. 2011. Computing Political Preference among Twitter Followers, *Proceedings of CHI* 2011.

Goldberg, LR. 1990. An alternative "description of personality": The Big-Five factor structure, *Journal of Personality and Social Psychology* 59(6): 1216-1229.

Gutmann, A and Thompson, D. 2004. *Why Deliberative Democracy?* Princeton, New Jersey: Princeton University Press.

Habermas, J. 1962. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society.* Translated by T Burger (1989). Cambridge, MA: MIT Press.

Habermas, J. 1984. *The Theory of Communicative Action.* Translated by Thomas McCarthy. Boston: Beacon Press.

Hall, MA and Smith, LA. 1999. Feature Selection for Machine Learning: Comparing a Correlation-based Filter Approach to the Wrapper, *Proceedings of the Twelfth International FLAIRS Conference*, AAAI.

Hansen, DL and Johnson, C. 2012. Veiled viral marketing: disseminating information on stigmatized illnesses via social networking sites, *Proceedings of International Health Informatics Symposium* (IHI) 2012.

Hart, W; Albarracín, D; Eagly, AH; Brechan, I; Lindberg, MJ; Merrill, Lisa. 2009. Feeling Validated versus Correct: A meta-analysis, *Psychological Bulletin* 135(4): 555-588.

Hinsz, VB; Tindale, RS; Vollrath, DA. 1997. The emerging conceptualization of groups as information processors, *Psychological Bulletin* 121(1): 43-64.

Holtgraves, TM. 1997. Yes, but...: Positive politeness in conversation arguments, *Journal of Language and Social Psychology* 16: 222-239.

Hong, L and Page, SE. 2001. Problem Solving by Heterogeneous Agents, *Journal of Economic Theory* 97(1): 123-163.

Horrigan, J; Garret, RK; Resnick, P. 2004. The Internet and Democratic Debate: Wired Americans hear more points of view. Pew Internet & American Life Project, Washington, DC.

Iyengar, S and Hahn, KS. 2009. Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use, *Journal of Communication* 59: 19-39.

Janssen, D and Kies, R. 2005. Online Forums and Deliberative Democracy, *Acta Politica* 40(3).

Jeh, G and Widom, J. 2002. SimRank: a measure of structural-context similarity, *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*: 538-543.

Jones, EE and Harris, VA. 1967. The attribution of attitudes, *Journal of Experimental Social Psychology* 3(1): 1–24.

Kahneman, D. 2011. *Thinking, Fast and Slow.* Farrar, Straus and Giroux.

Kastenmüller, A; Greitemeyer, T; Jonas, E; Fischer, P; Frey, D. 2010. Selective exposure: The impact of collectivism and individualism, *British Journal Of Social Psychology*, 49(4), 745-763.

Katz, E. 1957. The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis, *Public Opinion Quarterly* 21(1): 61-78.

Kelly, J; Smith, M; Fisher, D. 2005. Opinion Diversity in Online Political Discussion Networks, *Online Deliberation* 2005.

Kittur, A; Chi, EH; Suh, B. 2009. Crowdsourcing User Studies With Mechanical Turk, *Proceedings of CHI* 2009: 453-456.

Klapper, JT. 1960. *The effects of mass communication.* New York: Free Press.

Knobloch-Westerwick, S and Meng, J. 2009. Looking the Other Way: Selective Exposure to Attitude-Consistent and Counterattitudinal Political Information, *Communication Research* 36(3): 426-448.

Koop, R and Jansen, HJ. 2009. Political Blogs and Blogrolls in Canada: Forums for Democratic Deliberation? *Social Science Computer Review* 27(2): 155-173.

Kopfman, JE; Smith, SW; An Yun, JK, Hodges, A. 1998. Affective and cognitive reactions to narrative versus statistical evidence organ donation messages, *Journal of Applied Communication Research* 279-300.

Krebs, V. 2005. It's the Conversations, Stupid! The Link between Social Interaction and Political Choice, *In Extreme Democracy*, edited by J Lebkowsky and M Ratcliffe.

Krebs, V. 2008. Political Polarization in Amazon.com Book Purchases. orgnet.com.

Kriplean, T; Morgan, J; Freelon, D; Borning, A; Bennett, L. 2012. Supporting Reflective Public Thought with ConsiderIt, *Proceedings of CSCW* 2012.

Kriplean, T; Toomim, M; Morgan, J; Borning, A; Ko, AJ. Is this what you meant? Promotiing Listening on the Web with Reflect, *Proceedings of CHI* 2012.

Krosnick, JA and MacInnis B. 2010. Frequent Viewers of Fox News Are Less Likely to Accept Scientists' Views of Global Warming. A report of the Woods Institute for the Environment, Stanford University.

Kull, S; Clay, R; Lewis, E. 2003. Misperceptions, Media, and the Iraq War, *Political Science Quarterly* 118(4): 569-598.

Kullback, S and Leibler, RA. 1951. On information and sufficiency, *Annals of Mathematical Statistics* 22:79-86.

Kwak, H; Chu, H; Moon, S. 2011. Fragile online relationship: a first look at unfollow dynamics in Twitter, *Proceedings of CHI* 2011: 1091-1100.

Landis, JR and Koch, GG. 1977. The measurement of observer agreement for categorical data, *Biometrics* 33: 59-174.

Lawrence, E; Sides, J; Farrell, H. 2010. Self-Segregation or Deliberation? Blog Readership, Participation, and Polarization in American Politics, *Perspectives on Politics* 8: 141-157.

Lazarsfeld, PF; Berelson, B; and Gaudet, H. 1944. *The people's choice*, New York: Columbia University Press.

Li, I; Dey, A; Forlizi, J. 2010. A stage-based model of personal informatics systems, *Proceedings of CHI* 2010, 557-566.

Locke, EA and Latham, GP. 2002. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9): 705-717.

Locke, EA and Latham, GP. 2005. Goal Setting Theory. In KG Smith, & MA Hitt, *Great minds in management: the process of theory development* (pp. 128-150). Cambridge, UK: Oxford University Press.

Lowin, A. 1967. Approach and avoidance: alternative modes of selective exposure to information, *Journal of Personality and Social Psychology* 6(1): 1-9.

Lowin, A. 1969. Further evidence for an approach-avoidance interpretation of selective exposure, *Journal of experimental and social psychology* 5: 265-71.

MacKuen, M; Wolak, J; Keele, L; Marcus, GE. 2010. Civic Engagements: Resolute Partisanship or Reflective Deliberation, *American Journal of Political Science* 54(2): 440-458.

Mansbridge, J. 1980. *Beyond Adversary Democracy.* New York: Basic Books.

Mason, W and Watts, DJ. 2009. Financial incentives and the 'performance of crowds,' SIGKDD Workshop on Human Computation: 77-85.

Matlin, C; Singer-Vine, J; Wilson, C. 29 April 2010. Escape from the Echo Chamber, *Slate.* http://www.slate.com/id/2252247/.

McCallum, A and Nigam, K. 1998. A comparison of event models for Naive Bayes text classification, *AAAI-98 Workshop on Learning for Text Categorization.*

McKenna, L and Pole, A. 2007. What do bloggers do: an average day on an average political blog, *Public Choice* 134(1-2): 97-108.

Mei, Q; Guo, J; Radev, D. 2010. DivRank: the Interplay of Prestige and Diversity in Information Networks, *Proceedings of KDD* 2010: 1009-1018.

Menchen-Trevino, E and Karr, C. 2011. "Researching real-world Web use with Roxy: Collecting observational Web data with informed consent," *Journal of Information Technology & Politics Conference on The Future of Computational Social Science.*

Mill, JS. 1865. *Considerations on Representative Government.* London: Longman, Green, and Roberts.

Morris, JS. 2005. The Fox News Factor, *The International Journal of Press/Politics* 10(3): 56-79.

Morris, JS. 2007. Slanted Objectivity? Perceived Media Bias, Cable News Exposure, and Political Attitudes, *Social Science Quarterly* 88(3): 707-728.

Mullainathan, S and Shleifer, A. 2005. The market for news, *American Economic Review* 95: 1031–1053.

Munson, SA; Resnick, P. 2011. The Prevalence of Political Discourse in Non-Political Blogs, *Proceedings of ICWSM* 2011.

Munson, SA; Resnick, P. 2010. Presenting Diverse Political Opinions: How and How Much, *Proceedings of CHI* 2010.

Munson, SA; Zhou, DX; Resnick, P. 2009. Sidelines: An Algorithm for Increasing Diversity in News and Opinion Aggregators, *Proceedings of ICWSM* 2009.

Mutz, DC. 2006. *Hearing the Other Side: Deliberative versus Participatory Democracy.* New York, N.Y.: Cambridge University Press.

Mutz, DC and Martin, PS. 2001. Facilitating Communication Across Lines of Political Difference: The Role of Mass Media, *American Political Science Review* 95(1): 97-114.

Mutz, DC and Mondak JJ. 2006. The Workplace as a Context for Cross-Cutting Political Discourse, *The Journal of Politics* 68(1): 140-155.

Negroponte, N. 1995. *Being Digital.* New York: Alfred A Knopf.

Nemeth, CJ. 1986. Differential contributions to majority and minority influence, *Psychological Review* 93(1): 23-32.

Nemeth, CJ and Rogers J. 1996. Dissent and the search for information. *British Journal of Social Psychology* 35: 67-76.

Noelle-Neumann, E. 1993. *The spiral of silence: public opinion, our social skin.* 2nd edition. Chicago, Illinois: University of Chicago Press.

Ng, EW and Detenber, BH. 2005. The Impact of Synchronicity and Civility in Online Political Discussions on Perceptions and Intentions to Participate, *JCMC* 10(3).

Obama, B. 2010. Commencement Address at the University of Michigan, as transcribed by the *Ann Arbor Chronicle.* 1 May 2010.

Oh, A; Lee, H; Kim, Y. 2009. User Evaluation of a System for Classifying and Displaying Political Viewpoints of Weblogs, *Proceedings of ICWSM* 2009.

Papacharissi, Z. 2004. Democracy online: civility, politeness, and the democratic potential of online political discussion groups, *New Media and Society* 6(2): 259-283.

Pariser, E. 2011. *The Filter Bubble: What the Internet is Hiding from You.* Penguin Press.

Park, S; Kang, S; Chung, S; Song, J. 2009. NewsCube: delivering multiple aspects of news to mitigate media bias, *Proceedings of CHI* 2009.

Park, S; Ko, M; Kim, J; Liu, Y; Song, J. 2011. The Politics of Comments: Predicting Political Orientation of News Stories with Commenters' Sentiment Patterns, *Proceedings of CSCW* 2011.

Pew Research Center for the People and Press. 2004. *Online news audience larger, more diverse: News audience increasingly polarized.* Washington, DC: Pew Research Center.

Pew Research Center's Project for Excellence in Journalism. 2005. *The State of the News Media 2005.* Washington, DC: Pew Research Center.

Pew Research Center's Project for Excellence in Journalism. 2012. *The State of the News Media 2012: An Annual Report on American Journalism.* Washington, DC: Pew Research Center.

Price, V and Cappella, JN. 2002. Online Deliberation and Its Influence: The Electronic Dialogue Project in Campaign 2000, *IT & Society* 1(1): 303-329.

Prior, M. 2007. *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections,* Cambridge University Press.

Procaccia, AD; Rosenschein, JS; Zohar, A. 2008. On the complexity of achieving proportional representation. *Social Choice and Welfare, 30*(3): 353-362.

Purcell, K; Rainie, L; Mitchell, A; Rosentiel, T; Olmstead, K. 2010. *Understanding the participatory news consumer: How Internet and cell phone users have turned news into a social experience.* Pew Research Center, The Pew Internet & American Life Project, and The Pew Project for Excellence in Journalism.

Putnam RD. 2001. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Putnam RD. 1993. *Making Democracy Work: Civic Traditions in Modern Italy.* Princeton, New Jersey: Princeton University Press.

Ramsay, C; Kull, S; Lewis, E; Subias, S. 2010. Miscommunication and the 2010 Election: A Study of the US Electorate, WorldPublicOpinion.org Project, University of Maryland. http://drum.lib.umd.edu/handle/1903/11375

Rendall, S. 2001. Fox's Slanted Sources; Conservatives, Republicans far outnumber others, Fairness and Accuracy in Report.

Rosenburg, M. 1955. Some Determinants of Political Apathy. *Public Opinion Quarterly* 18 (Winterly): 349-366.

Ross, L. 1977. The intuitive psychologist and his shortcomings: Distortions in the attribution process, in *Advances in experimental Social Psychology,* edited by L. Berkowitz. New York, New York: Academic Press.

Ross, L; Greene, D; and House, P. 1977. The False Consensus Effect: An Egocentric Bias in Social Perception and Attribution Processes. *Journal of Experimental Social Psychology* 13: 279-301.

Sanders, GS and Mullen, B. 1982. Accuracy in perceptions of consensus: Differential tendencies of people with majority and minority positions. *European Journal of Social Psychology* 13(1): 57-70.

Schelling, TC. 1984. Self-Command in Practice, in Policy, and in a Theory of Rational Choice, *The American Economic Review* 74(2): 1-11.

Schkade, D; Sunstein, CR; Hastie, R. 2007. What Happened on Deliberation Day? *California Law Review* 95(3): 915-940.

Scheufele, DA; Nisbet, MC; Brossard, D; Nisbet, EC. 2004. Social Structure and Citizenship: Examining the Impacts of Social Setting, Network Heterogeneity, and Informational Variables on Political Participation, *Political Communication* 21(3): 315-338.

Schultz, PW; Nolan, JM; Cialdini, RB; Goldstein, NJ; Griskevicius, V. 2007. The Constructive, Deconstructive, and Reconstructive Power of Social Norms, *Psychological Science* 18(5): 429-434.

Sears, DO. 1965. Biased indoctrination and selectivity of exposure to new information, *Sociometry* 28: 363-376.

Sears, DO and Friedman, JL. 1967. Selective Exposure to Information: A Critical Review, *Public Opinion Quarterly* 31(2): 194-213.

Seijts, GH; Latham, GP; Tasa, K; Latham, BW. 2004. Goal Setting and Goal Orientation: An Integration of Two Different Yet Related Literatures. *Academy of Management Journal*, 47(2): 227-239.

Sheng, VS; Provost, F; Ipeirotis, PG. 2008. Get another label? improving data quality and data mining using multiple, noisy labelers, *Knowledge Discovery and Data Mining* 2008 (KDD-2008).

Sibona, C and Walczak, S. 2011. Unfriending on Facebook: Friend Request and Online/Offline Behavior Analysis, *Proceedings of HICSS* 2011: 1-10.

Smith, A. 2011. Twitter and Social Networking in the 2010 Midterm Elections. Washington, DC: Pew Internet & American Life Project.

Stromer-Galley, J. 2003. Diversity of Political Opinion on the Internet: Users' Perspectives. *Journal of Computer-Mediated Communication* 8(3).

Stucke, ME and Grunes, AP. 2011. Why more antitrust immunity for the newspaper industry is a bad idea, *Northwestern Law Review Colloquy* 105: 115-133.

Sukumaran, A; Vezich, S; McHugh, M; Nass, C. 2011. Normative influences on thoughtful online participation, *Proceedings of CHI* 2011: 3401-3410.

Sunstein, C. 2001. *Republic.com.* Princeton, New Jersey: Princeton University Press.

Sunstein, C. 2002. The Law of Group Polarization, *The Journal of Political Philosophy* 10(2): 175-195.

Swearingen, K and Sinha, R. 2002. Interaction Design for Recommender Systems, *Proceedings of DIS* 2002: 70-80.

Sweeney, PD and Gruber, KL. 1984. Selective exposure: voter information preferences and the Watergate affair, *Journal of Personality and Social Psychology* 46(6): 1208-21.

Tabor, CS and Lodge, M. 2006. Motivated Skepcisim in the Evaluation of Political Beliefs, *American Journal of Political Science* 50: 755-769.

de Tocqueville, A. *On Democracy in America*, volume 1 (1835) and volume 2 (1840). Translated by Henry Reeve, 1899. Vintage Books Edtion, 1990.

Tumasjan, A; Sprenger, TO; Sandner, PG; Welpe, IM. 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment, *Proceedings of ICWSM* 2010: 178-185.

Van Alstyne, M and Brynjolfsson, E. 1996. Electronic Communities: Global Village or Cyberbalkans?, *Proceedings of the International Conference on Information Systems.* Cleveland, Ohio.

Witten, IH and Frank, E. 2005. *Data Mining: Practical machine learning tools and techniques*, 2nd Edition. San Francisco: Morgan Kaufmann.

Wojcieszak, ME and Mutz, DC. 2009. Online Groups and Political Discourse: Do Online Discussion Spaces Facilitate Exposure to Political Disagreement?, *Journal of Communication* 59(1), 40-56.

Wojcieszak, ME; Baek, YM; Carpini, MXD. 2010. Deliberative and Participatory Democracy? Ideological Strength and the Processes Leading from Deliberation to Political Engagement, *Int J Public Opin Res* 22(2): 154-180.

Yano, T and Smith, NA. 2010. What's worthy of comment? Content and Comment Volume in Political Blogs, *Proceedings of ICWSM* 2010.

Zaller, JR. 1992. *The Nature and Origins of Mass Opinion.* Cambridge University Press.

Zhou, XD; Resnick, P; Mei, Q. 2011. Classifying the Political Leaning of News Articles and Users from User Votes, *Proceedings of ICWSM* 2011.

Zhou, XH; Obuchowski, NA; McClish, DK. 2002. *Statistical Methods in Diagnostic Medicine.* New York: Wiley & Sons.