

8-5-2011

# Mitigating Selective Filtering's Polarizing Effect on Web 2.0 Content

David Schuff

*Temple University*, david.schuff@temple.edu

Ozgur Turetken

*Ryerson University*, turetken@ryerson.ca

Robert D. St. Louis

*Arizona State University*, stlouis@asu.edu

Follow this and additional works at: [http://aisel.aisnet.org/amcis2011\\_submissions](http://aisel.aisnet.org/amcis2011_submissions)

---

## Recommended Citation

Schuff, David; Turetken, Ozgur; and St. Louis, Robert D., "Mitigating Selective Filtering's Polarizing Effect on Web 2.0 Content" (2011). *AMCIS 2011 Proceedings - All Submissions*. 137.

[http://aisel.aisnet.org/amcis2011\\_submissions/137](http://aisel.aisnet.org/amcis2011_submissions/137)

This material is brought to you by AIS Electronic Library (AISEL). It has been accepted for inclusion in AMCIS 2011 Proceedings - All Submissions by an authorized administrator of AIS Electronic Library (AISEL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Mitigating Selective Filtering's Polarizing Effect on Web 2.0 Content

**David Schuff**  
Temple University  
David.Schuff@temple.edu

**Ozgur Turetken**  
Ryerson University  
turetken@ryerson.ca

**Robert D. St. Louis**  
Arizona State University  
stlouis@asu.edu

## ABSTRACT

For almost two decades, the Internet and related technologies have made more information available to information users than they can handle. The decentralization of content creation that is a feature of Web 2.0 has only exacerbated this problem. This state of overload, combined with our tendency toward hypothesis-confirming behavior, can result in biased information selection, and threatens both civil discourse and effective decision-making. In this paper, we describe a study of a technique designed to mitigate filtering by enabling content consumers to see a greater diversity of information. The results of our experiment support the notion that the strength of people's opinions can be changed by reading relevant information, but provide only weak support for the effectiveness of categorizing information content. We discuss how the results will guide our future research and inform theory and practice.

## KEYWORDS

Web 2.0, blogosphere, information filtering, information overload, systems design

## INTRODUCTION

The first decade of the 21<sup>st</sup> century saw two parallel and connected trends in technology and the media. The first was a fundamental shift in the creation and distribution of information. With the advent of Web 2.0 technologies such as blogging, it became both easy and inexpensive to mass-distribute content; due to the falling price of computer hardware, the ubiquity of access to the Internet, and a generation of web-based tools that do not require technical skills. The second trend was increased fragmentation and polarization of our public discourse. In the United States, news organizations such as Fox News and MSNBC carry a point of view in order to appeal to a specific audience. Even news outlets that do not explicitly identify their orientation, such as The New York Times or the Wall Street Journal, frequently are identified as liberal or conservative. The strategy can be effective, as people consume media in line with their point of view (Iyengar and Hahn, 2009). This has accompanied an ongoing polarization of the American electorate, as districts become increasingly more conservative or liberal (Abramowitz, Alexander, and Gunning, 2006).

We propose that these trends, while they have emerged separately, are strongly related. The enormous volume of user generated content is far too much for anyone to realistically consume. As a result, people must employ filters in choosing information. In the absence of effective support mechanisms, these filters tend to be based on people's prior biases. This can be seen in the public discourse that takes place online. Most blogs, like the mainstream media, convey a specific point of view, populated by like-minded individuals. While most blogs do not claim to be unbiased, they are clearly a source of information and analysis for many people. One might consider the web a "Habermasian" abstract public sphere (Habermas, 1996), an open forum for discussion where people can share their ideas. However, it is still easy for blog readers to get caught within a "silo" of opinion, and be exposed only to a subset of potential viewpoints.

Adamic and Glance (2005) found that a blog's inbound links tend to come primarily from blogs with a similar orientation. Users with limited time and attention are likely to experience "information overload," a state where their need to process

information that merits attention is greater than their ability to process it (Schultze and Vandenbosch 1998). In this environment, they are likely to start with a like-minded, trusted source and stay within that locus of links. Alternatively, they may filter based on popularity, which is primarily determined by the number of inbound links a blog post receives.. Using popularity as a filtering criterion (i.e., the stories listed first on a Google Blog Search, or the “most popular” section on a specific blog) is particularly troublesome with Web 2.0 media, because it creates a self-perpetuating power-law scenario (Shirky, 2003). The sources with more inbound links are seen by more people, are shared more often, and therefore become increasingly popular. This makes it increasingly difficult for less-popular sources to gain traction. The dominance of a few sources is common in traditional media because there are relatively few newspapers and television channels, and they are not intended to provide a forum for deliberation. However, this runs contrary to the role of the public sphere that the blogosphere could potentially fulfill.

There is another foreseeable set of issues for organizations that embrace Web 2.0 content. Companies view crowdsourcing and user generated content from employees and customers as a way to gain new insights and ideas. They hope to effectively turn these internal and external constituents into a virtual “team of rivals,” (Sonnenfeld, 2002) benefitting from their collective knowledge and judgment. However, little prevents managerial decision-makers from being similarly overwhelmed with content, forcing them to filter. Unassisted, these choices are unlikely to be optimal, resulting in key perspectives being missed. One can easily imagine managers making information consumption choices based on popularity (frequently read posts), source, or similarity to their own viewpoint. Therefore, our research question is as follows:

*How does the presentation of issue-based Web 2.0 content shape users’ opinions?*

We explore this issue by demonstrating how manipulating the presentation of user-generated content can make it easier for people to expose themselves to multiple points of view. This should moderate their opinion, reduce polarization, and set the foundation for more reasoned decision-making. Building upon a prototype by Schuff, Turetken, and Asif (2010) and drawing from theory on the public sphere and human-computer interaction, we conduct a study as an initial test of our hypotheses and our study design. We discuss the results of our experiment, as well as the lessons learned.

## **HOW OPINIONS BECOME SILOED**

It is well-established that the human capacity for processing information is limited (Miller, 1956; Mandler, 1967). When faced with tasks that challenge these limitations, people employ coping mechanisms. Previous studies have found that when performing tasks that involve expending effort, people will try to minimize their “cognitive cost” through the use of decision aids (Benbasat and Todd, 1996). They also sacrifice outcome quality in favor of speed (Todd and Benbasat, 1992). People are inherently biased toward effort minimization.

We also know that people employ hypothesis confirming behavior when screening information to support their decision-making. Lowin (1967, 1969) found that people favor messages that enable them to reinforce their existing beliefs. They prefer either strong confirmatory messages, or “weakly dissonant” ones that can be easily refuted. They avoid compelling, “strongly dissonant” messages. Lowin examined this issue in the context of political communications, but people exhibit this behavior in other scenarios, including evaluating individual performance (Darley and Gross, 1983) and learning about new products (Hoch and Deighton, 1989). This has significant implications for decision-making, as we know people seek out confirmatory evidence to support their decisions (Frey, Schulz-Hardt, and Stahlberg, 1996). This effect becomes further heightened as they acquire and screen more information (Jonas, Schulz-Hardt, Frey, and Thelen, 2001).

Manjoo (2008) proposes that it is not only seeking consonant messages, but also seeking consonant sources that make this effect particularly powerful. He cites a study by Iyengar and Hahn (2009) that found people preferred news stories labeled as coming from sources that conformed to their political orientation, regardless of content. The combination of our tendency to engage in information selection based on source and content implies that we prefer to limit ourselves to a narrow band of opinion, despite what we claim about our desire for balanced sources.

We propose that this selection bias is exacerbated by people's inherent desire to reduce the cognitive cost of information consumption and decision-making. Hoch and Deighton (1989) suggest that people seek concordant messages, in part, because the cost of making an error is less than the cost of a full search. It follows from Benbasat and Todd (1996) that when faced with an overwhelming amount of information, people are even more likely to rely on decision-aids. Filtering discordant messages is not only hypothesis-confirming behavior, but also a coping mechanism. Therefore, people may be more willing to consume a greater diversity of information if the cost of consuming content is lowered. That is, it may be possible to mitigate hypothesis-confirming behavior by making filtering less necessary.

Habermas (1974, 1989) characterizes an abstract public sphere as a distributed arena of voices where all can participate. A functioning public sphere is essential for group decision making, and the blogosphere has a unique potential to fulfill that role. Filtering can severely limit that potential. Sunstein (2004) claims the problem of user-driven content filtering is so significant that it threatens public discourse and democracy. Cayzer (2004) argues that navigation tools are needed to assist users in navigating the blogosphere's growing content.

In the same way that societal discourse is threatened, this expanding body of Web 2.0 content can threaten the ability of organizations to deliberate. Managers are faced with a growing set of (often conflicting) information. They must make choices about what information will inform their decisions, and may limit themselves to a few "trusted" sources. Considering Frey et al.'s (1996) finding that people are predisposed to seek information confirming their decisions, it is likely that the flood of viewpoints contained within user generated content may actually *reduce* the diversity of information considered.

## TOWARDS A REMEDY

We propose that tools that effectively leverage user-generated content, such as the content in the blogosphere, must have two essential components. First, they must support navigation through vast amounts of content. Second, they must reduce the need for people to apply their own filters to cope with the overload. To define the fundamental attributes of such a tool, we consider O'Baioill's (2004) requirements for the blogosphere to serve as a functioning public sphere:

- *The blogosphere should be inclusive.* The reader should be exposed to as much information as possible, and information should not be systematically excluded.
- *Consumption of blog content should not be guided by rank.* In other words, popularity should not be used as a filtering mechanism.
- *There should be support for rational debate.* The structure of the blogosphere should support the exchange and discussion of ideas.

The explosion of user-generated Web 2.0 content undermines these basic principles. The process of filtering is often biased and therefore threatens inclusivity. Many of the mechanisms for navigation, such as blog rolls and ranked search, are based on popularity and therefore violate the principle of disregard of rank. The application of filtering in this way compromises the legitimacy of communication, making it difficult to engage in a rational debate of ideas.

In our earlier work (Schuff et al., 2010), we developed a set of design principles for navigation tools to support O' Baoill's principles. The goal was to guide the creation of tools that minimize bias while also reducing overload. The result was a prototype tool based on clustering the text of a set of blog posts. The tool organized the posts by content, with visual cues used to aid navigation within and between the clusters. We build on that work by updating the design and testing its effectiveness in shaping user opinions.

Text clustering has great potential as a remedy for the problems with biased filtering. Van Rijsbergen's cluster hypothesis (1979) states that documents similar in content serve similar information needs. People perform better on cognitive tasks when information is organized into groups (Mandler, 1967). Meaning is conveyed through the groupings, and this additional information makes it easier to understand the associations between them (Ashcraft, 1989). This ultimately enables people to

reduce their cognitive load and process more information by moving between chunks of data, instead of navigating a flat set of information all at once.

There is another important aspect of text clustering in the context of the blogosphere. This method of organization groups documents based purely on the words each document contains. It inherently disregards source, rank, and point of view because the clustering algorithm only looks at word frequencies across documents. The category titles become unbiased filters that readers can use to find the topics they are looking for, and blog posts with differing points of view and popularity should be combined within a single cluster.

Therefore, clustering is an important way to address O' Baoill's (2004) principles. Since the clustering process disregards source and allows people to consume more information by reducing overload, it supports inclusivity. Source popularity is ignored in the clustering process, facilitating disregard of rank. Enabling people to consume information from a greater diversity of sources should support rational debate.

## USER STUDY

### The Prototype Tool

The prototype tool for this study is an updated version of the initial design implemented in Schuff et al. (2010). While the design of this prototype (see Figure 1) differs significantly from that version, both applications are built around the notion of facilitating navigation through the clustered results of a collection of blog posts.

The application displays the output generated by the open-source clustering engine Carrot2. The input for Carrot2 is an XML document containing the content of the blog posts to be clustered. The clustering engine generates a second XML file describing the title and content of each cluster. Our web application combines these two files to create a navigable display of clusters and posts.

The user navigates the collection of blog posts by first selecting a topic from the list on the left side of the display. The topics on the list are taken from the cluster titles generated by Carrot2. The right side of the display is an accordion-style component that shows the detail of the selected cluster by listing the blog posts contained within it. When an individual blog post is selected, the entry expands to display the text of the post, while the other blog post titles remain visible.

Several visual cues are employed to facilitate navigation. These cues can be extracted "preattentively" (Crapo, Waisel, Wallace, and Willemain, 2000), allowing the user's attention to remain on the content. We use size to indicate the number of blog posts within a particular topic. Topics containing more blog posts are displayed in a larger font. This can be used as an indication of the relative importance of a topic, or the relative level of activity within it. Color is used to provide continuity between the topic list and the cluster detail. The selected cluster is highlighted, allowing the user to view the other cluster titles without losing track of where they are within that list.

We propose that the blog clustering tool should lower the cognitive cost of consuming a greater diversity of information. First, organizing information by topic, instead of source, increases comprehensibility (Cukier, Ngwenyama, Bauer and Middleton, 2009) by making it easier to investigate multiple points of view. Second, by ignoring popularity (i.e., inbound links), legitimacy (Cukier et al., 2009) of the information set is increased, making it more likely that clusters will be composed of a heterogeneous mixture of sources. Third, the user interface is designed to aid navigation across and within topics, reducing the user's overload. This encourages them to read more, and more relevant, information.

We first seek to establish the role of reading blogs in changing opinion, regardless of the tool used. This is an important indicator of the potential to break down the "siloing" of opinion as it addresses whether opinion *can* be changed. We can measure this change by examining the number of subjects whose answers to the survey questions change when they consume blog content relevant to the target topic. Therefore, we hypothesize:

*H1: Subjects who read a collection of blogs relevant to the target topic will more frequently change their strength of belief than those who read a collection of non-relevant blogs.*

**Figure 1: The Prototype Application Used in This Study**

The exposure to more points of view is designed to break down the siloing of opinion. This can be measured by observing the degree to which the treatment makes the user's opinion regarding an issue more moderate. It has been argued that people's positions on issues get more extreme if they are exposed only to points of view that are similar to their own (Sunstein, 2006). Therefore, we use the moderation of people's positions as a surrogate for exposure to a more diverse and larger amount of material as well as an outcome in itself. Accordingly, we hypothesize:

*H2: Reading from a collection of blogs organized using the clustering tool will moderate opinions more than reading from the same collection presented as a list.*

## Experimental Procedure

We conducted a controlled laboratory experiment to test our treatments and hypotheses. The subjects were 79 undergraduate and graduate students from a large University in Toronto, Ontario. The study consisted of two separate sessions; subjects who completed both sessions were given \$25 CDN and entered in a drawing to win an Apple iPad.

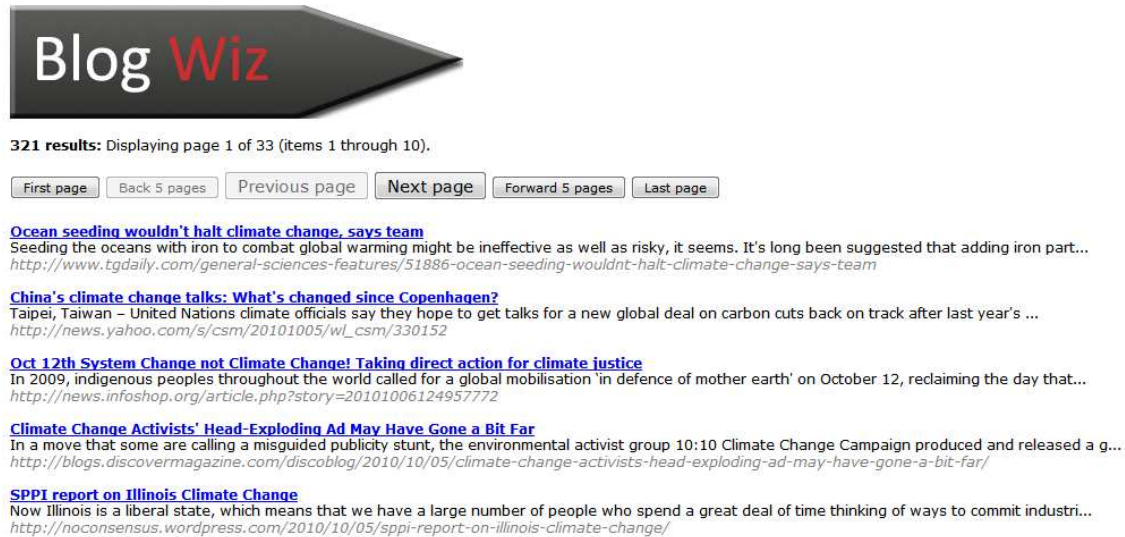
During the first session, which lasted about 30 minutes, subjects provided basic demographic information such as age, gender, and level of Internet and web use. Then they were asked to respond to 15 statements (see Table 1) regarding the strength of their beliefs about global warming (from 1=strongly disagree to 7=strongly agree). We chose this topic because there is a range of opinion on this issue in the blogosphere, as well as among individual citizens.

The second session was held two weeks later, so subjects would not recall exactly how they originally answered the questions. They were given instructions for using the tool and told that they had 45 minutes to read the blog posts in any order, although they may not be able to read all of them in the allotted time.

Item	Statement
1	I am knowledgeable about global warming.
2	I have strong opinions about global warming.
3	I am confident about my opinion on global warming.
4	I am concerned about significant changes to the climate occurring within a generation.
5	The causes of climate change are well-understood.
6	The changes in the climate have been overstated in the popular press.
7	Developing nations should be exempt from international climate change treaties.
8	Global warming is mostly caused by people.
9	Potential negative impacts to the economy outweigh the benefits from reducing people's impact on the environment.
10	The effects of global warming can be reduced or reversed by changes in people's behavior.
11	I would buy energy only from "clean" sources, even if they are more expensive.
12	The government should increase taxes on individuals to subsidize and develop alternative sources of energy.
13	Reducing carbon emissions places an unreasonable burden on businesses and the economy.
14	I would give up my car and take public transportation in order to lower my impact on the environment.
15	Corporations should be legally required to reduce their energy consumption, even if it increases their cost of doing business.

**Table 1. Statements Given to Subjects**

The subjects were assigned to one of three groups. The first group was given 321 blog posts about global warming to read using the prototype blog reader described in the previous section (see Figure 1). The blog posts were taken from 244 blogs with various points of view. The second group was given the same posts, but the tool was a simple ordered-list interface (see Figure 2). The third group was given the same interface as the second group, but was given a set of 97 posts from unrelated celebrity and entertainment blogs. These three groups allowed us to examine and control for the effects of both the presentation method and the content.



**Blog Wiz**

321 results: Displaying page 1 of 33 (items 1 through 10).

[First page](#)
[Back 5 pages](#)
[Previous page](#)
[Next page](#)
[Forward 5 pages](#)
[Last page](#)

**[Ocean seeding wouldn't halt climate change, says team](#)**  
 Seeding the oceans with iron to combat global warming might be ineffective as well as risky, it seems. It's long been suggested that adding iron part...  
<http://www.tgdaily.com/general-sciences-features/51886-ocean-seeding-wouldnt-halt-climate-change-says-team>

**[China's climate change talks: What's changed since Copenhagen?](#)**  
 Taipei, Taiwan – United Nations climate officials say they hope to get talks for a new global deal on carbon cuts back on track after last year's ...  
[http://news.yahoo.com/s/csm/20101005/wl\\_csm/330152](http://news.yahoo.com/s/csm/20101005/wl_csm/330152)

**[Oct 12th System Change not Climate Change! Taking direct action for climate justice](#)**  
 In 2009, indigenous peoples throughout the world called for a global mobilisation 'in defence of mother earth' on October 12, reclaiming the day that...  
<http://news.infoshop.org/article.php?story=20101006124957772>

**[Climate Change Activists' Head-Exploding Ad May Have Gone a Bit Far](#)**  
 In a move that some are calling a misguided publicity stunt, the environmental activist group 10:10 Climate Change Campaign produced and released a g...  
<http://blogs.discovermagazine.com/discoblog/2010/10/05/climate-change-activists-head-exploding-ad-may-have-gone-a-bit-far/>

**[SPPI report on Illinois Climate Change](#)**  
 Now Illinois is a liberal state, which means that we have a large number of people who spend a great deal of time thinking of ways to commit industri...  
<http://noconsensus.wordpress.com/2010/10/05/sppi-report-on-illinois-climate-change/>

**Figure 2: Ordered-list Treatment with Global Warming Blog Posts**

At the end of the session, subjects again responded to the 15 statements about global warming. The statements were presented in a different order than the first session to further ensure that subjects would not simply remember their original answers and repeat them.

## DATA ANALYSIS

To test H1, we compared the number of times subjects in the first two treatment groups (who were given global warming blogs) changed their strength of belief to the number of times subjects in the third treatment group (who were given entertainment blogs) changed. A change in strength of belief was coded as a binary variable, with a “1” indicating a subject moved two or more points on the scale, and a “0” if they moved less than two points. We selected two points as the cutoff to rule out “noise” between the two sessions, as it is reasonable to assume people may not respond completely consistently over time even if their strength of belief does not change.

The results of this analysis are listed in Table 2. We found that in 12 of the 15 statements, the subjects in groups 1 and 2 changed their strength of belief more often than subjects in group 3. We then conducted a binomial test to determine the likelihood of that outcome based on the assumption that there was no effect (i.e., both groups show a change in opinion at the same rate). The probability of observing 12 out of 15 “changes” given that  $\pi = 0.5$  is  $p < 0.0178$ . Therefore, H1 was supported.

We tested H2 by computing two measures of moderation. The measure for the first three statements (concerning the confidence in one’s opinion) was computed as a continuous variable representing the change between the two sessions, and can range from -6 to +6. A positive number indicates an increase in the subject’s overall level of confidence, and a negative number represents a decreased level of confidence. We equate a decreased level of confidence in one’s opinion as a more moderate view, likely a result of exposure to more diverse points of view.

For the remaining twelve statements, we computed a continuous variable representing the degree of moderation in strength of belief between the two sessions. For each subject, the moderation score was computed as follows:

$$\text{modscore} = |(session2Response - 4)| - |(session1Response - 4)|$$

This resulting “moderation score” can range from -3 to +3. Negative values indicate a movement toward the center, or a moderation of one’s viewpoint, and positive values indicate movement away from the center, or that the subject’s opinion became more extreme.



Questionnaire Item (Statement)	% of Groups 1 and 2 that moved 2+ points	% of Group 3 that moved 2+ points	Change in Group 1 > Change in Group 2
1	5.4%	0.0%	Yes
2	10.7%	0.0%	Yes
3	14.3%	4.3%	Yes
4	7.1%	4.3%	Yes
5	12.5%	4.3%	Yes
6	12.5%	8.7%	Yes
7	10.7%	8.7%	Yes
8	1.8%	0.0%	Yes
9	16.1%	8.7%	Yes
10	12.5%	4.3%	Yes
11	14.3%	8.7%	Yes
12	16.1%	21.7%	No
13	8.9%	17.4%	No
14	23.2%	26.1%	No
15	19.6%	13.0%	Yes

**Table 2. Comparison of Change in Opinion across Treatment Groups (H1)**

An ANOVA was conducted to test for differences between treatment groups (see Table 3). H2 was generally not supported, although there were a few significant differences. For example, in the case of questions 3 and 15 (confidence in one's opinion and whether corporations should reduce energy consumption), those assigned to the clustered condition became more moderate, while those in the other two groups became more extreme. However, the difference between groups 1 and 2 was not significant, while the difference between groups 1 and 3 (the control) was significant ( $p < 0.091$  and  $p < 0.064$ ). For question 4 (concern over climate changes), group 1 expressed less extreme opinions while group 2 became more extreme ( $p < 0.015$ ), although group 1 did not differ significantly from the control group. For question 8 (global warming is caused by people), groups 1 and 2 expressed less extreme opinions while group 3 became more extreme, although only the difference between groups 1 and 3 was significant ( $p < 0.049$ ). Note, however, that for all of the instances where there was a statistically significant difference, the subjects in group 1 (the subjects who used the clustering tool) became more moderate.

Questionnaire Item (Statement)	Group 1 Change Mean (sd) n=29	Group 2 Change Mean (sd) n=27	Group 3 Change Mean (sd) n=23	p-value (1 versus 2)	p-value (1 versus 3)
1	-0.17 (1.136)	-0.07 (0.730)	-0.22 (0.518)	0.704	0.861
2	-0.10 (1.145)	0.26 (1.059)	-0.04 (0.706)	0.225	0.827
3	-0.31 (1.628)	0.26 (1.130)	0.35 (0.935)	0.137	<b>0.091</b>
4	-0.31 (0.891)	0.30 (0.912)	-0.09 (0.900)	<b>0.015</b>	0.376
5	0.10 (1.145)	0.07 (1.035)	0.13 (0.920)	0.920	0.927
6	-0.21 (1.320)	0.00 (1.038)	0.17 (0.887)	0.519	0.241
7	-0.21 (0.940)	-0.41 (1.010)	-0.04 (1.147)	0.445	0.575
8	-0.24 (0.739)	-0.33 (0.832)	0.22 (0.902)	0.663	<b>0.049</b>
9	0.03 (1.149)	-0.07 (1.174)	-0.30 (0.559)	0.728	0.201
10	0.03 (0.944)	-0.04 (0.980)	-0.09 (0.793)	0.782	0.624
11	0.21 (0.861)	0.00 (1.074)	0.39 (1.076)	0.428	0.495
12	0.07 (0.998)	-0.11 (1.086)	0.17 (1.114)	0.521	0.722
13	0.00 (0.964)	-0.19 (0.879)	0.22 (0.795)	0.457	0.388
14	-0.24 (1.683)	0.41 (1.500)	-0.17 (1.072)	0.135	0.868
15	-0.07 (1.033)	0.19 (1.210)	0.43 (0.843)	0.401	<b>0.064</b>

**Table 3. Analysis of Moderation of Opinion across Treatments (H2)**

Note also that group 1 exhibited more of a movement toward moderation (or at least less of a movement toward an extreme) than group 3 for 12 of the 15 statements. A binomial test of this result given  $\pi = 0.5$  is  $p < 0.018$ . Similarly, group 2 exhibited more of a movement toward moderation (or at least less of a movement toward an extreme) for only 9 of the 15 statements. This result has a  $p < 0.304$  (and therefore is not statistically significant). This indicates that participants in group 1 were more likely to exhibit moderation of belief than the control group, but participants in group 2 may not have been.

## **DISCUSSION AND NEXT STEPS**

While we only found weak support in our experiment for hypothesis 2, there were some encouraging results. First, in support of H1, subjects did appear open to changing their strength of belief when given relevant information. This indicates the potential to manipulate the strength of subjects' opinions, even given the relatively short amount of time they were given to read the blogs (45 minutes).

What remains an open question is whether the method of presentation affects how opinions are influenced. There are encouraging results from the test of H2 – in 9 of the 15 items, subjects using the clustered presentation expressed a more moderate stance in the second session than the first. However, in 7 out of the 15 items, subjects given the ordered list presentation also became more moderate. Further, we found some evidence that the clustered presentation is more likely to yield more moderate opinions than the ordered list condition when compared to the control group. While the lack of statistically significant differences between the two treatment groups prevents any strong assessment of these findings, our results can guide the design of future studies. Below we identify four areas on which to focus.

### **Increase the Sample Size**

One possible explanation for the weak statistical support for Hypothesis 2 is that the study lacked sufficient statistical power. Running the experiment on a larger group of people is likely to resolve this issue. If the findings consistent with our hypotheses become significant while the others remain insignificant, this would confirm that a lack of power was the cause.

### **Consider Alternative Clustered Presentations**

Another possibility is that the categories generated by the clustering engine were not meaningful enough to direct subjects to the information in which they were interested. The computer generated categorization may have been no more helpful than an ordered list, which would not have reduced the cognitive cost enough to cause a difference in filtering behavior.

A remedy for this would be, for experimental purposes, to create our own clusters. This would enable us to guarantee semantically meaningful labels, and ensure that the labels matched the content of the categorizations. The issue with this is one of practicality – this would not be a practical way to implement a map of a large set of blog posts. However, it would enable us to isolate the effect of presentation without confounding it with cluster quality.

### **Examine Clickstream Data**

An underlying assumption of this study is that an improved presentation method will increase exposure to more diverse points of view. We indirectly measure this through the change in extremeness of people's opinions. However, we can measure this directly through clickstream data collected during the experimental sessions. This would allow us to see exactly which blog posts were selected, and would give us a better understanding of subjects' behavior. For example, we may find that subjects were reading a greater diversity of information in the clustering condition, but were still disregarding messages in conflict with their existing beliefs. It may be that being better informed, while a good outcome in itself, doesn't result in significant changes in opinion.

## Incorporate Personality Data

The relevance of the target issue to the individual, and the individual's general willingness to see other points of view could play a moderating role in the effectiveness of the presentation method. It may be that subjects who are more open to changing their mind may be influenced more by the clustered treatment, since it increases their exposure to multiple points of view. Integrating data from personality assessment instruments may provide more nuanced insight into how presentation method influences the extremity of opinion.

## CONCLUSIONS

The explosion of user-generated content will continue as Web 2.0 technologies move even further into the mainstream. In this study, we found preliminary evidence that blog content can influence people's strength of belief and tested a mechanism for clustering content that may mitigate the polarizing effect of selective filtering. The results from this line of research inform both theory and practice. Improving our understanding of how people filter when faced with large amounts of conflicting information will enable us to better guide the design of tools to support societal discourse and information sharing. Organizations that seek to leverage the "crowd" need ways of effectively digesting and integrating this information into their decision-making process.

## ACKNOWLEDGMENTS

This study was partially supported by a discovery grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada. We would like to thank Ashneet Gujral for his work in programming the prototype used in this study. We would also like to thank Sarah Madden and Charles Segal for their assistance in executing the experiment.

## REFERENCES

1. Abramowitz, A.I., Alexander, B., and Gunning, M. (2006) Incumbency, Redistricting, and the Decline of Competition in U.S. House Elections, *Journal of Politics*,68,1,75-88.
2. Adamic, L.A. and Glance, N. (2005) The political blogosphere and the 2004 U.S. election: Divided they blog, in *Proceedings of the Third International Workshop on Link Discovery*, New York, ACM Press,36-43.
3. Ashcraft, M.H. (1989) Human memory and cognition, Scott, Foresman and Co., Glenview, IL.
4. Benbasat, I. and Todd, P. (1996) The effects of decision support and task contingencies on model formulation: a cognitive perspective, *Decision Support Systems*,17,4,241-252.
5. Cayzer, S. (2004) Semantic blogging and decentralized knowledge management, *Communications of the ACM*,47,12,47-52.
6. Crapo, A.W., Waisel, L.B. , Wallace, W.A., and Willemain, T.R. (2000) Visualization and the process of modeling: A cognitive-theoretic view," in *Proceedings of the International Conference on Knowledge Discovery and Data Mining*, August 20-23, New York, ACM Press, 218-226.
7. Cukier, W., Ngwenyama, O., Bauer, R. and Middleton, C. (2009) A critical analysis of media discourse on information technology: preliminary results of a proposed method for critical discourse analysis, *Information Systems Journal*,19,175-196.
8. Darley, J.M. and Gross, P.H. (1983) A hypothesis-confirming bias in labeling effects, *The Journal of Personality and Social Psychology*,44,1,20-33.
9. Frey, D., Schulz-Hardt, S., and Stahlberg, D. (1996) Information seeking among individuals and groups and possible consequences for decision making in business and politics, in E. Witte and J. Davis (Eds.), *Understanding group behavior Vol. II: Small group processes and interpersonal relations*,Mahwah,NJ, Erlbaum,211-225.
10. Habermas, J. (1974) The public sphere: An encyclopedia article, *New German Critique*,3,49-55.
11. Habermas, J. (1989) The structural transformation of the public sphere: An inquiry into a category of bourgeois society. MIT Press, Cambridge,MA.

12. Habermas, J. (1996) *Between facts and norms: Contributions to a discourse theory of law and democracy*, MIT Press, Cambridge, MA.
13. Habermas, J. (2006) Political communication in media and society – Does democracy still enjoy an epistemic dimension? The impact of normative theory on empirical research, *Communication Theory*,16,4,411-426.
14. Hoch, S.T. and Deighton, J. (1989) Managing what consumers learn from experience, *The Journal of Marketing*,53,2,1-20.
15. Iyengar, S. and Hahn, K.S. (2009) Red media, Blue media: Evidence of ideological selectivity in media use, *Journal of Communication*,59,1,19-39.
16. Jonas, E., Schulz-Hardt, S., Frey, D., and Thelen, N. (2001) Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information, *Journal of Personality and Social Psychology*,80,4,557-571.
17. Lowin, A. (1967) Approach and avoidance: Alternate models of selective exposure to information, *Journal of Personality and Social Psychology*,6,1,1-9.
18. Lowin, A. (1969) Further evidence for an approach-avoidance interpretation of selective exposure, *Journal of Experimental Social Psychology*,5,3,265-271.
19. Mandler, G. (1967) Organization in memory, in K.W. Spence and J.T. Spence (Eds.), *The Psychology of Learning and Motivation*, Academic Press, New York, NY,327.
20. Manjoo, F. (2008) *True enough: Learning to live in a post-fact society*, Wiley, Hoboken.
21. Miller, G.A. (1962) Information input overload, in M.C. Yovits, G.T. Jacobi, and G.D. Goldstein (Eds.), *Proceedings of the Conference on Self-Organizing Systems*, Spartan Books, Washington.
22. Ó Baoill, A. (2004) Weblogs and the public sphere, in L.J. Gurak , L. Antonijevic, C. Johnson, C. Ratcliff, and J. Reyman (Eds.), *Into the blogosphere: Rhetoric, community, and culture of weblogs*. Retrieved February 19, 2011, [http://blog.lib.umn.edu/blogosphere/weblogs\\_and\\_the\\_public\\_sphere.html](http://blog.lib.umn.edu/blogosphere/weblogs_and_the_public_sphere.html).
23. Schuff, D., Turetken, O., and Asif, Z. (2010) Designing systems that support the blogosphere as a public sphere, *AIS Transactions on Computer Human Interaction*,2,3,95-111.
24. Shirky, C. (2003) Power laws, weblogs, and inequality, retrieved February 19, 2011, [http://www.shirky.com/writings/powerlaw\\_weblog.html](http://www.shirky.com/writings/powerlaw_weblog.html).
25. Sonnenfeld, J.A. (2002) What makes great boards great, *Harvard Business Review*,80,9,106–113.
26. Sunstein, C.R. (2004) Democracy and filtering, *Communications of the ACM*,47,12,57-59.
27. Sunstein, C.R. (2006) *Infotopia: How many minds produce knowledge*, Oxford University Press, USA.
28. Todd, P. and Benbasat, I. (1992) The use of information in decision making: an experimental investigation of the impact of computer based decision aids, *MIS Quarterly*,16,3,373–393.
29. van Rijsbergen, C. (1979) *Information retrieval*, 2<sup>nd</sup> Edition, Butterworths, London.