Citizen 2.0:

Public and Governmental Interaction through Web 2.0 Technologies

Kathryn Kloby
Monmouth University, USA

Maria J. D'Agostino
John Jay College of Criminal Justice, CUNY, USA

Information Science
REFERENCE

Chapter 2

An Experiment in E-Rulemaking with Natural Language Processing and Democratic Deliberation

Peter Muhlberger
Texas Tech University, USA

Jenny Stromer-Galley
University at Albany, USA

Nick Webb
University at Albany, USA

ABSTRACT

Public comment processes in federal and state agency rulemakings are among the most substantial potential arenas for public input into government. Unfortunately, these processes have not been much used for thoughtful public input. This chapter explores whether online democratic deliberation and natural language processing tools, can empower participants to provide more informed input into an agency rulemaking. It also sought to determine whether such an approach had other positive effects such as enhancing citizenship and increasing confidence in the pertinent agency. Findings indicate improvements in participant knowledge of the network neutrality rulemaking topic, systematic attitude change, improvements to citizenship measures, and increased confidence in the Federal Communications Commission. Results suggest that public deliberation under conditions needed to involve substantial numbers of people—namely, online deliberation without facilitators—can improve public comments into federal and state agency rulemakings while strengthening the citizenship qualities of participants. They also indicate that many of the desired effects of face-to-face deliberation with trained facilitators can also be obtained online without facilitators.

INTRODUCTION

E-government researchers and practitioners are increasingly interested in using online technologies not merely for transaction processing but also to open a public space for citizen-to-citizen and citizen-to-government interactions. One of the greatest potential spaces for meaningful public input into government is the domain of public comments on government agency proposed rules. Yet, research suggests that public comments tend to be of relatively low quality when it comes to their utility for government agencies (Shulman, 2006). Public and even expert comments in the rulemaking process might be improved by utilizing deliberative techniques that involve small group discussion. Held online, such discussion would permit large numbers of geographically dispersed individuals to participate. Deliberation is, however, typically conducted with trained facilitators. Agencies would be hard pressed to find trained facilitators for even a moderate number of small groups and face concerns about accusations of bias against these facilitators. A possible solution would be to combine Natural Language Processing (NLP) technologies to serve some of the functions of facilitators: answering questions, summarizing discussion, connecting people with common interests, eliciting comments from the quiet, and so forth. Thus, our project tested a combination of NLP technologies that would cover facilitator functions, in the hope of building a Discussion Facilitation Agent (DiFA).

This chapter describes the NLP technologies used and presents research from an experiment on the policy topic of network neutrality, which is the question of whether rules should be adopted to require internet service providers to treat all digital traffic equally. Our research question is whether online and unfacilitated deliberations and NLP technologies would lead to some of the positive outcomes of deliberation expected from theory and prior research on face-to-face deliberation with facilitation. The experiment showed that deliberation had impacts on a range of outcomes—such as increased topic knowledge, attitude change, positive changes in indicators of citizen engagement, and increased confidence in the FCC—but that the technologies deployed had limited effects. The chapter concludes with an examination of the limitations of the study and potential ways in which the NLP technologies could be further developed to enhance citizen deliberation on policy in the future.

E-DEMOCRACY AND RULEMAKING

The recent trend in open government initiatives in the United States and elsewhere has brought about noteworthy new opportunities for citizen engagement not only in interfacing with government but in participating more deeply in the policy process. The Obama administration's efforts on open government led to three principles: transparency, collaboration, and participation ("Transparency and Open Government," 2009). One of the ways that open government initiatives can manifest is through using the myriad of applications cast under the umbrella of "Web 2.0" (Chun, Shulman, Sandoval, Hovy, 2010). These technologies primarily include social media that enable collaboration and sharing among individuals, helping to facilitate the "wisdom of crowds" (Shirky, 2008). ICTs in general have been looked to for increasing efficiency and inter-agency cooperation as well as interfacing more effectively with the public (Margetts, 2009), in what Dunleavy, Margetts, Bastow and Tinkler (2006) refer to as a new "Digital Era of Governance".

Of particular relevance for the project at hand are the opportunities for increased participation by citizens in the policy process. Although there have been some limited efforts to engage the public in federal policy making (Heidinger, Buchmann, & Böhm, 2010), one of the underutilized and understudied opportunities is through the public comment process in government agency rulemakings.

Such rulemaking is the process by which agency officials seek to turn the often general language of legislation into specific rules that determine compliance with a law. Legislatures, such as Congress, make laws that are often abstract rather than specific guidelines. Agencies need to implement these laws by providing specific guidelines, that is rules. Before putting such rules in place, agencies typically seek public comment on proposed rules over a period of months or years. At the federal level, the Administrative Procedures Act (APA), which is the law governing how federal agencies can propose and enact regulations, requires the elicitation and consideration of public comments for significant rulemakings. The APA's Section 553 on rule makings stipulates that regulations proposed by government agencies must be published in the Federal Register, and section c explains, "After notice required by this section, the agency shall give interested persons an opportunity to participate in the rule making through submission of written data, views, or arguments with or without opportunity for oral presentation." The law also requires that the agency consider the "relevant matter presented." In brief, agencies must place proposed rules in the Federal Register, solicit public comments on them, and make a good faith effort to consider all comments. Agencies are motivated to follow this APA procedural format because if they do not, their rules will likely be challenged successfully in court. Because of their extensive use in federal and in many state agencies and their impact on actual rules enforced by government, the rulemaking comment process represents perhaps the most extensive arena within which the public could directly influence public policy. Rulemaking has great potential for substantive democratic engagement, though it is also an arena of which much of the public is unaware.

Although research suggests that the public comments agencies receive about a rulemaking have shaped final regulations (Yackee, 2005), there has been ongoing concern about the nature and content of public comments (Shulman, 2006), and the especially limited participation by citizens in providing comments (Kerwin, 2003). When citizens do involve themselves in the public comment process it is more commonly through the submission of form letters prompted by activist or lobbying groups of which citizens are members. In the efforts to motivate members to send comments, activist organizations frame the public comment process as a plebiscite, majority rule system where comments are perceived almost as votes (Shulman, 2009). Some research has examined the characteristic differences in form letters by activist groups versus original comments drafted by stakeholders, noting that original comments tend to provide more of the kinds of insight government agencies seek (Scholsberg, Zavestoski, & Shulman, 2007).

The advent of e-rulemaking in the U.S., through Regulations.gov, a one-stop portal for citizens to review regulations open for comment, and submit their comments electronically, led to optimism for increased citizen-involvement (Carlitz & Gunn, 2005). As e-mail campaigns and digital form letters have become more prevalent in the past decade, researchers have expressed concern that mass form letters are of limited utility for agencies seeking insight on the regulation (Shulman, 2009). On the other side of the debate is the position that the tactic by activist groups to mobilize members to send form letters is part of a complex public communication campaign and does not crowd out more original comments (Karpf, 2010). Other research suggests that the concern of interest or activist groups deluging agencies with public comments, either in the older print or newer digital forms, is misplaced. Instead, industry and stakeholder elites have been and continue to be significantly more likely to submit comments on environmental regulations over the past decade (de Figueiredo, 2006; Shafie, 2008). Put another way, the initial hope that moving the public comment process online would increase citizen involvement has not come to pass.

DEMOCRATIC DELIBERATION

Whether the public is swamping government agencies with meaningless form letters or is vastly outstripped in many rulemakings by elites, the public's overall impact on rulemaking outcomes is small at best. This raises concerns about the democratic legitimacy of agency rulemakings. While impartial experts should have final say on factual matters, in a democracy legitimacy for the values chosen in a policy decision comes from the public (Benhabib, 1994). A flourishing body of new work in deliberative democracy theory provides a compelling case for consulting the *informed* views of the public in making policy decisions (Bohman, 1996; Chambers, 1996; Dryzek, 2002; Gutmann & Thompson, 1996).

Public deliberation practice seeks to elicit more informed views of the public through a variety of methods. One widely used method is the Deliberative Poll® (Fishkin, 1997), in which a random sample of the public is recruited to participate in a one or two day deliberative event. Participants are provided with balanced information on the policy issue to read prior to the event, discuss the issue face-to-face in small groups of about 10 people with a trained facilitator, and question experts. Participants are also typically surveyed prior to and after discussion, allowing for pre- to post-discussion comparisons to determine the impact of deliberation on policy views, knowledge, and indicators of future engagement. The objective of using a representative sample of the public in a Deliberative Poll® is to determine what the general public would think about an issue if it were to collectively learn about and thoughtfully discuss the issue. While not permitted to call their work Deliberative Polls®, a number of researchers and deliberation practitioners utilize deliberation designs bearing some similarity.

Research on deliberation has arrived at a number of moderately firm empirical conclusions. Deliberations enhance the knowledge of participants with respect to the policies under discussion

and generally result in significant and substantial changes in participant attitudes (Barabas, 2004; Farrar et al., 2010; Luskin, Fishkin, & Jowell, 2002). Somewhat more tentatively established, deliberation results in enhanced legitimacy for government policy decisions (Zillig, Herian, Abdel-Monem, Hamm, & Tomkins, 2010) and shifts in attitudes and predispositions that suggest greater propensity for future political engagement (Gastil, Black, Deess, & Leighter, 2008; Price & Cappella, 2002). Generally, these findings regard pre-discussion to post-discussion changes, with researchers only beginning to unpack what features of deliberation result in the changes (Muhlberger & Weber, 2006c; Stromer-Galley & Muhlberger, 2009d).

Importantly for this project, little or no research has been conducted to determine whether the positive changes resulting from deliberation can be obtained without trained facilitators. Also, much deliberation research has proceeded without an experimental control group. Arguably, in a one or two-day face-to-face deliberation there need not be a great concern with such threats to internal validity as history (outside factors, such as current events that may cause observed pre- to post-discussion changes) or maturation (changes in participants, such as growing cognitive complexity with maturation, that can cause pre- to post-discussion changes). For an overview of experimental and quasi-experimental methods and threats to internal validity see Campbell, Stanley, and Gage (1966). Because we introduce a control group here, the current study involves a higher standard of evidence than much deliberative research. Finally, comparatively little peer-reviewed literature exists regarding the impacts of online deliberation (Muhlberger & Weber, 2006c; Price & Cappella, 2002).

If deliberation is to become a more widely used process or even an everyday process in government, such as rulemaking, it will need to scale up. Such scaling is improbable for face-to-face deliberations with geographically dispersed

groups because of the considerable cost of such events. In addition, as already noted, any substantial number of discussion groups would require a rare commodity in the Deliberative Polling® paradigm—trained facilitators. Thus, there is a need for research, such as presented here, on online deliberation in the absence of facilitators.

THE DELIBERATIVE E-RULEMAKING PROJECT

Natural Language Technology

The Deliberative E-Rulemaking (DeER) project was a three-year National Science Foundation project which sought to apply deliberation and NLP tools to federal and state agency rulemakings. The study described in the current paper is one of multiple efforts under DeER to test such e-rulemakings. Because federal and state agencies were ultimately not especially cooperative with the DeER project, the current data represents the largest body of data collected in the project. The technology deployed in the DeER project was chosen to address specific problems of multi-user deliberation among citizens. One problem is uninformed participants. Making available large amounts of contextual documentation that provides background research on the topic at hand places a large burden on the user. Some of the material may be already known to participants, while other information may be irrelevant to their concerns. Moreover, having vested parties to the deliberation put the documentation together can lead to concerns of bias in the material. Instead, using existing question-answering technology, we allowed users to explore the policy issue using natural language queries, resulting in paragraphs of information drawn from a pre-mined document base from the World Wide Web, that provided information relevant to the user and the deliberation.

We made use of the HITIQA (High-Quality Interactive Question-Answering) tool (Small &

Strzalkowski, 2009), developed at the University at Albany. HITIQA helps users find answers to complex analytical problems. That is, this system is not a factoid QA system, or one that returns solely single-fact answers to direct questions (such as "How tall is the Eiffel Tower?"), but rather is designed to answer difficult, compound questions ("What is the impact of the falling dollar on oil operations in the Caspian Sea?") by facilitating exploration through a series of returned paragraphs or subsets of complete documents. HITIQA uses the process of text framing to bring a level of semantic representation to open-domain data in order to facilitate meaningful dialogue with the user. In experiments using intelligence analysts, we demonstrated that HITIQA is more effective for research-based applications than using Google. Originally written with professional analysts in mind, HITIQA was designed with interactive features to help the user narrow extracted answers through a series of questions. Such use was unlikely for citizen users in a Web-based environment. Thus, we disabled the interactive capabilities of HITIQA, and focused on its use as a one-question retrieval mechanism for relevant data, with results presented in a list ordered by relevance.

A significant element of the evaluation process for HITIQA in the DeER project was to observe if users, with minimal training, would pose natural language queries to the HITIQA engine at all. There is a plethora of evidence (cf. HitWise report for 2010) that users are accustomed to posing short keyword queries. The average number of keywords to Google is between 2 and 3. In this environment, and told that HITIQA could answer their questions, how would users interact with the system?

This application of HITIQA was a 'pull' technology; it required users to be motivated to search the background documents. As the DeER project progressed, it became clear that few users were motivated to pose questions to the HITIQA system, preferring instead to opt for familiar search

engines. This may partly be a function of having not highly motivated users or of users unfamiliar with NLP (something that will change with time). What was required was a 'push' technology that would present information more actively to the user. This involved answering non-rhetorical questions in individual postings *during a deliberation* and linking the question text to the result.

To determine which questions were non-rhetorical and meaningful and how to best answer them, it was necessary to classify discussion posting sentences. The technology we used is called dialogue act (DA) classification. Dialogue acts are simple labels applied to sentences that indicate the function each sentence is playing in the wider discourse. For example, we want to be able to differentiate statements of opinion from questions. Within questions, it is also useful to distinguish between wh-questions (who, what, where and so on), yes-no questions, ("Do you…?"), and rhetorical questions (which expect no particular answer). We employed a state-of-the-art dialogue act classifier (Nick Webb, Hepple, & Wilks, 2005), which is applicable to a wide range of dialogue types (Webb & Ferguson, 2010).

We applied DA classification to users' posts, with two aims. First, we automatically identified questions in the text, and directed them to the HITIQA system automatically. In the background, HITIQA retrieved paragraphs of data from our database, and linked them to the identified question. So, if a user asked, in a post "How many government agencies regulate the national parks?," such a question would be identified, sent to HITIQA, and then the material HITIQA retrieved would be linked to the question text in the post in which it occurred. Clicking this link opened a pop-up answer page containing the returned paragraphs from HITIQA. Thus, users did not have to actively explore the contextual material.

Second, we used DAs as a base mechanism to link users together. Another hypothesis of deliberation is that, once an active deliberation on an asynchronous message board (that is, not real-time—posts are stored and people can participate at any time they prefer) of the type we were using becomes too large, it is difficult for users to track new messages in the expanding set of threads. We used DA classification to identify clear statements of personal opinion in posts, and then used keyword overlap of those utterances to automatically identify people expressing opinions on the same or similar topics. At this point, we did not pay attention to the valence of those arguments. What mattered was connecting users based on similarity of interests.

To encourage participation in the deliberation, regular emails were automatically created which contained links to posts with overlapping perspectives, indicating new postings in threads other than those in which the targeted user was active, in which topics similar to the user's postings could be found. In addition, to give all users some understanding of the development of the overall, on-going deliberation, we included a word cloud (automatically created from wordle.net), which shows the most frequently used words (we filtered typical stop words) by size, the larger being the more frequent.

Ultimately, for both citizen and agency consumption, we would want to make available large scale, multi-posting summarization systems. Initially, we had intended to employ such a summarization system (based on our existing, multi-document summarizer system, XDOX (Hardy et al., 2002)), but required a substantial body of textual data to tune the system, which we were unable to obtain.

The NLP technologies offered to users were a small step toward building a discussion facilitation agent (DiFA). Like a human facilitator, the system sought to answer factual questions and sought to connect the ideas of participants when these participants addressed similar issues but at different points in the discussion. The word cloud also served to summarize the main ideas in a discussion and rapidly refer the participant to the contents of each idea. The system can be enhanced

in a variety of ways with existing technology, a point to which we return in the discussion.

Multi-Level Deliberation

The project utilized a Multi-Level Deliberation (MuLD) technique with a number of potential advantages, as suggested in prior theory and practice (Endenburg, 1998; Pivato, 2007). In MuLD, discussion groups eventually select one or a few persons to represent the group in a higher-level discussion group in which each member represents a lower-level group. In MuLD, participants discuss an issue in a small group—up to 30 people in an online setting. This group then elects, say, one of its members to represent the group in a second-level group of up to 15 people. Each member of that group represents a lower-level group. This second-level group discusses the issue, seeking to bring together the best ideas and reasoning from all the lower-level groups. If there are multiple second-level groups, each would elect a member for a third-level group. The process ends when there is one "top-level" group that represents all first-level discussants. Representatives of higher-level groups are also expected to return to their lower-level groups to explain what was decided in the higher-level groups and why.

It is possible for very large numbers of first-level discussants to share information efficiently among themselves in such a MuLD arrangement. In the example arrangement above, four levels of groups could encompass millions of discussants. Hypothetically, more informed and deliberative participants would flow to higher-level groups, improving the quality of discussion. The top-level group can, potentially, directly interact with policymakers as representatives of all discussants. Higher-level groups could utilize lower-level groups for such tasks as information processing, and potentially interact with agency officials in a meaningful way. Representatives relay the gist of their upper-level discussion back to their original groups, hopefully spreading good ideas from any group to all other groups and helping to legitimize higher-level decisions. MuLD could be crossed in many ways with online technologies to potentially improve communication, representation, and perceived legitimacy.

METHODS

Research Design

We conducted a partial 2X2 experimental manipulation—testing three of four crossed conditions, as depicted in Table 1. All discussion groups were online. Six of the discussion groups received access to our DiFA technology, including the question answering tool, and the email reminders with links to content-similar posts and a word cloud. These are depicted under "Technology Enabled" in Table 1. Another five groups were given only the opportunity to deliberate on an asynchronous message board, but without the additional tools. The number of groups per condition was unequal because of constraints to maintain a critical discussion mass in each group and uncertainty about the ultimate number of discussants. We also had three of the technology-enabled groups participate in a multiple level deliberation (MuLD), where they selected people to represent them in a higher-level group. The deliberations spanned a month.

Table 1. Research design—Number of discussion groups in each experimental condition

	Technology Enabled	Not Technology Enabled
Multiple Level Deliberation	3	
No Multiple Level Deliberation	3	5

Students in the MuLD condition were invited to vote for representatives two weeks into the deliberation. We did not test the fourth condition, with MuLD but no technology, out of concerns regarding sample size and, thereby, statistical power. This research can therefore clarify what impact MuLD has on top of or in interaction with technology, but not without technology.

We recruited 615 students from two public universities who participated in exchange for extra credit in several courses in communication. Students were randomly assigned to 11 online discussion groups that were hosted through an asynchronous, threaded message board system called VBulletin. Of the students that were recruited, 121 met the requirements for participation, including posting to the message board and answering pre- and post-discussion surveys. Another 63 students took the sign-up survey, but failed to take the pre-discussion survey and failed to participate in the discussions, but did take the "post-discussion" survey when allowed to do so for partial extra credit. These students served as a comparison group to determine whether changes observed among the discussion participants were really caused by the discussion and not some outside factors. We expected students in departments of communication to have at least a mild interest in the topic of network neutrality, which was the topic of the deliberation, and hoped that this interest would blossom as they discussed the issue. On the other hand, end of semester pressures might have dampened this possibility. In addition, it is unlikely that students would be as interested in the topic as the engaged public that would come to an agency rulemaking.

Network neutrality was presented to participants as the real issue of whether the Federal Communications Commission (FCC) should adopt a rule requiring all U.S. internet service providers (ISPs) to treat all digital traffic on their networks equally. Participants were told that proponents of the neutrality rule contend that without such a rule ISPs could throttle or block digital traffic in ways that would undermine free speech rights. Opponents, in contrast, contend that a network neutrality rule is unnecessary to protect free speech, but that such a rule would cause real problems for ISPs trying to offer such services as online phone calls or in protecting the internet access of average ISP user from the small number of people who use far more bandwidth. The network neutrality issue was selected because it was an important real issue at the time, with which the FCC was grappling, and because of its relevance for communications students. Participants were given access to a website with a "learn" page containing summaries of different pro and con arguments and links to the sources of these arguments. They also received access to our question-answering tool.

Measures

Participants answered two Web-based surveys: one before the deliberation began and another shortly after the deliberation concluded. The surveys were taken from any computer at the respondent's leisure. Because respondents were answering questions to a computer and without researchers present, responses should be relatively free of social desirability and demand effects. Survey items included standard demographic items and a wide variety of explanatory and outcome variables.

Direct Evaluation Measures

We adapted Schweiger, Sandberg, and Ragan's (1986) evaluation scales for work groups, measured with 7-point Likert-type scales. Participants were asked about their Satisfaction with Discussion (e.g., "I am committed to my discussion group's recommendations on net neutrality."), Future Participation Motivation ("I am motivated to participate in future public policy discussions."), and Reevaluated Assumptions ("Participating in this net neutrality policy discussion made me critically reevaluate the validity of the assumptions and recommendations that I held personally.").

Each of these three dimensions were measured with two questions and responses were averaged. Correlations within dimension were strong, and prior work establishes the presence of distinct factors (Stromer-Galley & Muhlberger, 2009d).

Respondents were also asked three questions regarding the perceived legitimacy of the deliberation and of FCC decisions. The variable "Impact Should Have On Officials" was measured with a question asking respondents to indicate how strongly FCC officials should weigh comments and conclusions from their discussion, with 0 meaning "have no effect on...the Officials' Decisions," 3 meaning "have moderate effect on...," and 6 meaning "determine...the Officials' Decisions." The "Impact On Officials – No Agree" variable asks much the same question but tells respondents to consider how much of an effect they would want on FCC officials' decisions if they personally disagreed with the comments and conclusions of other participants. The responses to the first and second question should clarify just how legitimate the deliberation process seemed to participants. The "Respect FCC" variable asks respondents to "Suppose the FCC does not end up taking the positions you do on net neutrality. If so, how strongly do you agree or disagree with the following statement: I will respect the FCC\'s decisions." The difference between the second and this third question should clarify how legitimate participants view FCC departures from their preferences in the absence of deliberation relative to deliberations that yield results different than the respondent would like.

A series of questions asked about the usefulness of such technological features as email links, the word cloud, and question answering. The basic structure of the questions was: "The 'Your Questions Answered' feature of the DeER website was useful." followed by a seven-point Likert scale. To measure the effectiveness of MuLD, two questions styled after the Schweiger measures asked respondents whether MuLD stimulated their thinking or uncovered valid questions or assumptions they had not considered.

Outcome Measures

Policy Knowledge was measured as the summed number of correct answers to a seven question quiz about network neutrality. For each quiz question, respondents needed to select one of four possible answer options. An example question is: "What are 'common carriers'?" (Correct answer: "private companies that are required by law to carry information without discrimination"). Policy Attitude was measured as an average of three questions, such as: "Internet Service Providers (ISPs) should treat all information on the Internet the same." Higher scores indicate greater acceptance of network neutrality. Active Citizenship, which might also be considered a measure of deliberative conceptions of citizenship, was captured with four questions such as: "A good citizen should allow others to challenge their political beliefs." Active Citizenship involves a conception of good citizenship that includes willingness to allow others to challenge one's views, justifying views to others who disagree, and willingness to listen to and discuss politics with others who disagree. Prior research establishes active citizenship as a distinct factor (Muhlberger & Weber, 2006c), indicates that it is correlated with self-reported past participation, and shows that people with an active conception of citizenship acquire more objective knowledge during the course of a deliberation. In addition, three studies indicate that active citizenship grows in the course of deliberation (Muhlberger, 2007). Citizen Identity measures the centrality of politics to personal identity with such questions as: "Being politically conscious is important to who I am." Such Citizen Identity is strongly related to self-reported past participation (Muhlberger, 2005a).

A substantial literature suggests that right-wing authoritarianism (RWA) adversely affects citizenship (Muhlberger, 2011b). People high in RWA are less knowledgeable about important social

facts, less politically active, and embrace forms of governance that minimize democratic input. RWA is measured with a shortened four-item scale with good factor properties (Muhlberger, 2011b), including such questions as "Obedience and respect for authority are the most important virtues children should learn." FCC Confidence was measured with an adapted government confidence question, "I approve of the way the Federal Communication Commission has been handling its job lately." Procedural Justice was captured with two questions adapted from Tyler's (2006) work, such as, "I have confidence that the Federal Communication Commission will take seriously the concerns of people like me."

RESULTS

We first explore the data to understand how engaged student participants were in the deliberation. This helps clarify expectations regarding the presence and size of deliberation effects, particularly relative to the engaged public of actual rulemakings. Second, we examine participant evaluations of the deliberation, the legitimacy of deliberative and FCC policy decisions, and evaluations of the DiFA technologies. These evaluations help clarify how participants experienced the deliberation, including their perceptions of deliberative quality. Third, we present results regarding the comparability of the comparison and experimental groups. With greater comparability, the comparison group can be relied upon to rule out the history and maturation threats to internal validity. Fourth, we present findings regarding pre- to post-discussion changes in policy knowledge, policy attitudes, and various indicators of engagement and citizenship, which may be strengthened via the deliberation. Finally, we consider whether participants who received more votes in the MuLD condition were more capable or engaged.

Engagement

One important consideration to put the current results into context is the degree to which participants in this study were engaged with the deliberation process. If students were highly engaged with the topic, we might expect to find more substantial effects of the deliberation, but, if not, weak or undetectable effects might be expected. Ideally, participants would either come to a deliberation interested in the topic or develop an interest during the discussion process.

A good measure of how engaged students were with the deliberation topic is how many times they posted. They were told that they would get class credit for participating, but only if they posted at least five thoughtful posts responsive to others on the bulletin board. Thus, a student seeking to do the bare minimum of work for the extra credit would post five times or perhaps a little more to provide some cushion in case the researchers decided that some of the posts did not qualify as thoughtful. We will examine only those students who posted two or more times because these students will be considered the group that deliberated in subsequent analyses. Participants who never posted cannot be said to have deliberated, and we will ignore those who posted only once because they constitute a gray zone.

Among participants who posted two or more times, 69% posted five or fewer times, thus either achieving the bare minimum requirement or less. Eighty-six percent posted six or fewer times, and 93% posted seven or fewer times. The vast preponderance of students, then, did not do more than the minimum amount of work, plus a small margin of one or two extra posts, likely to help insure they would receive extra credit. Few of the participants were sufficiently engaged in the process to suggest self-sustaining interest. This leads us to anticipate weak effects.

Table 2. Averages of participant evaluations of deliberation and of technologies, with t-tests

Variable	Mean (*s.d.* of mean)	Scale Range
Satisfaction with Discussion	1.00*** (.10)	-3 to 3
Future Participation Motivation	.77*** (.12)	-3 to 3
Reevaluated Assumptions	1.22*** (.10)	-3 to 3
Impact Should Have On Officials	3.79*** (.11)	0 to 6
Impact On Officials – No Agree	3.54*** (.10)	0 to 6
Respect FCC	.22 (.13)	-3 to 3
Email Links	.94*** (.18)	-3 to 3
Word Cloud	.65*** (.17)	-3 to 3
Question Answer System	.62*** (.16)	-3 to 3
Multi-Level Deliberation	.33* (.14)	-3 to 3

Notes: *N*=121 for all non-technology evaluation questions. *n*=63 for Email Links, Word Cloud, and Question Answer System, and *n*=38 for Multi-Level Deliberation. The mean is the mean value for each variable across all participants. *t*-tests are used to determine whether this mean is significantly different than zero—that is, whether respondents have significantly positive evaluations.

* is *p*<.05, two-sided; *** is *p*<.001, two-sided based on single-sample t-tests.

Evaluations

A weak but not ignorable form of evidence regarding the value of deliberation is participants' own evaluations of the experience. Table 2 shows a variety of evaluations participants in this project gave of the deliberations, the legitimacy of the deliberations, and of the technologies used. Included are the mean evaluation, standard deviation of the mean, information on whether the evaluation was significantly different than zero, and the scale of the variable. The full sample of 121 was used for all non-technology evaluations (other than MuLD). The technology evaluations could only be considered for those who were in the technology conditions, cutting the sample to 63. MuLD was tested on only 38 participants, so the numbers reflect only their responses.

The deliberation experience received modestly positive evaluations that are highly significantly above zero. General satisfaction and perceived reevaluation of assumptions due to the deliberation hover at or above one on a -3 to 3 scale. Motivation to participate in future deliberations averages to a modest but positive value (.77). The legitimacy of the proceedings appears to be quite substantial. Respondents believe that government officials should take into account the deliberation's conclusions about the network neutrality at a moderate level (3.79), where 0 means the officials should ignore the deliberation results and 6 means official choices should be determined by the deliberation results. A second legitimacy question, in which the respondent was instructed to imagine that they personally disagreed with the deliberation results still shows only a slightly diminished legitimacy level. When asked whether they would respect an FCC decision opposed to their own wishes on the net neutrality issue, respondents average a small value (.22) on a -3 to 3 Likert scale—not significantly different than zero. While the scale is not strictly comparable to the legitimacy scales, it does appear the respondents privilege the outcome of a deliberation, even one that takes a position in opposition to their preferences, above that of an FCC decision contrary to their preferences.

Respondents evaluated the technologies and the MuLD deliberation design from modestly favorable to very modestly favorable, but nonetheless positively and quite significantly so. Email links to possible posts of interest were viewed

most favorably, followed by the word cloud and QA system, and, least favored (but still positively favored), the MuLD deliberation design. The order of preference might be in terms of familiarity in prior experience.

Comparability of the Comparison Group

This study includes, as described earlier, a non-experimental comparison group—students who completed the first survey but never posted to the bulletin board. These students were asked to complete the post-discussion survey. While this comparison group is not a perfect control group, because it is self-selected and conceivably some of its members were exposed to some of the experimental manipulation, as online readings or exposure to messages about the discussion, they nevertheless may help build a case against some threats to internal validity.

To address the concern of self-selection, it is helpful to determine to what extent this non-discussant comparison group is self-selected. Does it differ significantly from discussants? In t-tests of differences on an array of variables, there were no significant or trend differences on: gender, ethnicity (Caucasian / non-Caucasian), need for cognition, citizen identity, or confidence in the FCC. There were non-significant trend differences (.05< p <.10) on network neutrality attitude and knowledge, with discussants showing mildly more opposition to network neutrality (-.34 vs. -.06 on a 7-point scale, a difference of 4% of the scale range) and mildly more knowledge about the issue (3.74 vs. 3.41 on an 8-point scale). There were two significant differences. Discussants were modestly further along in school (p=.02, two-sided; 3.22 vs. 2.81, where 1=freshman, 2= sophomore, etc.) and had slightly higher levels of active citizenship (p=.03; 1.54 vs. 1.27 on a 7-point scale or 3.9% of the scale range). Besides the possibility that some of these differences may be due to chance variation across a substantial number of indica-

tors, it is implausible that such small differences between these two groups could account for significant outcome differences between the groups. For example, active citizenship might account for somewhat more learning among a split-half of people high in active citizenship compared to the half low in this quality, but the impact these differences have on learning is small even where both groups are exposed to an intensive learning environment (Muhlberger & Weber, 2006c). With only a 4% scale difference in active citizenship between discussants and non-discussants in the current study, it seems unlikely that the observed differences would noticeably affect outcomes.

Deliberation Outcomes

Table 3 shows the impact of the deliberation conditions on the average *change* in various outcome variables from the pre- to post-discussion surveys. Three discussion conditions were tested: no-technology and no MuLD, technology and no MuLD, and technology and MuLD. Due to limited sample size and the group nature of the data, MuLD was added on top of the technology condition rather than tested both with and without technology. The intercept indicates the average change in the outcome variable ("Change in Variable") among discussants in the no technology and no MuLD condition. Subsequent coefficients indicate what needs to be added to the intercept to obtain the average change of the outcome variable in that column's condition. For example, the average change in policy knowledge for non-discussants is.34-.43=-.09—or effectively zero. Also, discussants in the technology and MuLD condition had an average change of citizen identity of.22-.33+.41=.30. Ideally, the intercept adjustment for non-discussants would be significant and of the opposite sign of the intercept (average change due to deliberation). This would indicate that discussants and non-discussants have significantly different average changes, suggesting that the deliberation manipulation had an effect that cannot

Table 3. Ordinary least squares regressions of post- minus pre- outcomes on deliberation conditions with cluster-robust standard errors

Change In Variable	Intercept (Avg. Change Due to Deliberation)	Intercept Adjust for Non-Discussants	Intercept Adjust for Technology Exposure	Intercept Adjust for Multi-Level Deliberation	$N; R^2; s.e.$
	Coefficient (standard error)				
Policy Knowledge	.34* (.17)	-.43* (.21)	-.03 (.23)	.15 (.22)	184;.044; 1.46
Policy Attitude	-.58*** (.11)	.22† (.17)	.13 (.15)	-.16 (.27)	184;.146; 1.22
Active Citizenship	.21** (.08)	-.34** (.13)	.04 (.15)	-.05 (.15)	184;.040;.93
Citizen Identity	.22** (.07)	-.03 (.15)	-.33* (.15)	.41** (.14)	184;.057;.90
Right-Wing Authorit.	.05 (.05)	-.12 (.08)	.05 (.08)	-30** (.09)	184;.035;.70
FCC Confidence	.29* (.13)	.05 (.19)	-.02 (.19)	-.03 (.19)	184;.027; 1.20
Procedural Justice	.11 (.14)	.09 (.21)	.02 (.19)	.10 (.19)	184;.054; 1.25

Notes: N=184 (deliberators plus comparison groups, N=121 and 63, respectively). Each row represents an OLS regression of the "Change in Variable" indicated in the left column on the conditions of deliberation listed in the remaining columns. The "Change in Variable" represents the difference in the variable from the pre-discussion to the post-discussion survey (post minus pre). The intercept is the constant for the model but also represents the average change in the outcome variable for participants in the no technology exposure and no multi-level deliberation condition. The remaining independent variables indicate what would have to be added to the intercept to obtain the average for that condition or combination of conditions. Reported R^2 values are corrected to include the variance explained by the intercept, which is an explanatory variable in this case. Each row of the table represents an OLS regression with the "Change in Variable" as the dependent variable and the remaining columns as independent variables. † is $p<.10$; * is $p<.05$; ** is $p<.01$, *** is $p<.001$. All p-values are two-sided, except for the non-discussant coefficient, which is expected to be in the opposite direction from the discussant coefficient and is therefore reported as one-sided. Errors are cluster robust (accounting for covariation of errors within discussion groups) and were estimated in R.

be dismissed as due to threats to internal validity such as history or maturation—for example, news events between the pre- and post-discussion surveys that may have resulted in a significant change in responses (history) or maturation of participants in this time period resulting in changes (maturation). This represents a higher standard of evidence than has been typical of deliberation research—that is, it is more difficult to establish both a change in the experimental group and no change in the comparison group. Because of imperfections of the comparison group—we cannot rule out that the group did not experience some of the experimental manipulations—it is possible for there to be real changes that will not be identified by this high standard.

Policy knowledge and active citizenship show the ideal pattern. Both show significant changes from pre- to post-surveys among discussants and also significantly less change for the comparison group. This provides strong evidence of an effect not clouded by potential internal validity threats.

Policy attitudes also significantly changed and the change for the comparison group was less but only shows trend significance. Citizen identity shows a significant improvement for discussants, though the value also appears to improve for the comparison group.

In general, this deliberation shows significant changes similar to those found in many face-to-face deliberations, particularly knowledge increases, attitude change, and some evidence for enhanced future citizenship, here measured as active citizenship and citizen identity. In addition, in contrast to literature critical of deliberation that predicts deliberation will undermine confidence in government, the evidence here regarding confidence (in the FCC) and procedural justice (the FCC respects and takes seriously the views of people like the respondent) indicates that these variables either show no change or significant improvement.

Technology exposure and exposure to MuLD does not increase objective knowledge nor change

policy attitudes, contrary to expectations. Such exposure does appear to have some impact on variables that seek to capture future engagement and citizenship potential. Technology exposure significantly reduces citizen identity. Exposure to MuLD significantly counteracts the effect of technology on citizen identity and significantly decreases right-wing authoritarianism, an effect that may be beneficial for citizenship.

The Best and the Brightest

One of the hopes for MuLD is that the representatives elected by group members to represent the group at higher levels will be more knowledgeable and capable than the average group member. In effect, MuLD would allow leaders to emerge from among discussants and raise the level of discussion. Alternatively, critics might claim that those elected will overrepresent the already socially dominant. To test these hypotheses in the current study, ordered probit analysis was used to regress the number of votes each discussant received on the discussant's internal and external efficacy, network neutrality policy knowledge, active citizenship, citizen identity, need for cognition, political interest and demographics (university class level, gender, ethnicity, and family income). None of these variables show a significant impact on number of votes, even at trend levels of significance ($p<.10$).

DISCUSSION

This project sought to address a little studied issue—the feasibility of technology-enabled deliberations as regular rather than one-off processes in government. In particular, we tested the use of Natural Language Processing (NLP) technologies in the context of citizen deliberation of a federal rulemaking issue—whether the Federal Communication Commission should adopt a network neutrality rule requiring Internet Service

Providers to treat all internet traffic equally. The federal and state agency rulemaking processes, which typically require public comment by law, promise a substantial, if underused, arena for direct public input into government in the U.S. We tested a deliberation under conditions that would likely prevail were such deliberations to become regular parts of the rulemaking process involving substantial numbers of geographically dispersed participants—namely, online discussion without facilitators, who are costly and unlikely to be available in substantial quantity. Such a test expands the meager research on online deliberation and the largely non-existent literature on deliberation in the absence of trained facilitators. Our hope was that a combination of NLP technologies, forming a Discussion Facilitation Agent (DiFA), would cover some of the functions of facilitators.

The results provide suggestive evidence for the value of online deliberation in the absence of facilitation. Participants gave significantly positive evaluations of the deliberation and its quality as well as of the NLP technologies tested (Table 2). With respect to legitimacy, participants believed that officials should moderately take into account the comments and outcomes of this deliberation—even supposing they personally did not agree with those comments and outcomes. On the other hand, they indicated little or no respect for unilateral FCC decisions running contrary to their personal views. For government agencies struggling to make decisions the public perceives as legitimate, this finding suggests some value in agencies using deliberative fora to enhance decision legitimacy. An interesting question for future research is whether members of the public who hear about an agency-held deliberation, rather than participating in it directly, will perceive the agency decision as more legitimate.

Importantly, this study finds significant pre- to post-discussion changes in a variety of deliberation outcome indicators that have been taken as the hallmarks of successful face-to-face deliberations with facilitation in prior research (Table 3).

Policy knowledge increases and policy attitudes systematically change. Two indicators of enhanced citizenship and likely future participation, Active Citizenship and Citizen Identity, show significant improvement. Also confidence in FCC decisions increases.

The current study improves on much past research by including a comparison group of participants who completed pre- and post- surveys but failed to post to the discussion. This non-discussant comparison group does not differ from the discussant experimental group in most respects. The two significant differences between the groups are small in size. On the other hand, it is conceivable that non-discussants were exposed to components of the experimental manipulation. They may have read some background materials on the topic of discussion or viewed some of the discussion. They were exposed to the "minimal manipulation" of being asked to consider themselves as part of a citizen deliberation on the issue. Still, the non-discussant group diverged significantly or nearly significantly from the experimental group on three of the five significant changes in outcomes. For policy knowledge and active citizenship, the non-discussant group showed significantly less change than the discussant group. For policy attitude change, the discussant group showed trend significance for less change than the discussant group. For these outcomes, then, there is some reassurance that history and maturation threats to internal validity likely can be discounted.

With respect to two other variables, Citizen Identity and FCC Confidence, the comparison group did not significantly differ in pre- to post-discussion change from the experimental group. Conceivably, the observed change in the comparison group might be wrought by the "minimal manipulation"—simply contemplating themselves in the role of citizen deliberator may have resulted in changes in both experimental and control groups. This, plus the merely trend significance on policy attitude change, indicate that

deliberation research may benefit from utilizing comparison or even control groups.

These significant outcome changes as well as positive evaluations are encouraging for the prospects for online deliberation in rulemaking. Nonetheless, it must be noted that, while positive, evaluations were not exuberant and outcome changes were small. Possibly, being online rather than face-to-face reduced the quality of the experience. Looming perhaps more heavily over this project was the low enthusiasm of the student participants, who showed little sign of self-sustaining interest in the deliberations beyond that required for receiving course extra credit.

The citizens who might volunteer to participate in an online deliberation in a real agency rulemaking would likely be more engaged with the issue of discussion. On the other hand, there remains the question of whether these citizens would take to online discussion as readily as college students. The longer-term trend, however, is for the public to become increasingly facile in the use of online technologies. More broadly, the citizens who participate in agency rulemakings are typically middle-class and well-educated, which suggests that voluntary rulemaking participants may not be less responsive, for reasons of education, to deliberation than our subjects. They may, instead, be rather more motivated because they self-select for interest in the rulemaking topic.

Certainly, researchers should seek to test online deliberation in real rulemakings. We sought to do so, but encountered widespread reluctance from multiple federal and state agency offices as well as more than 100 public interest groups—we suspect because citizen-based politics threatens established power brokers and power relationships (Muhlberger, Stromer-Galley, & Nick Webb, 2011). Perhaps the results of the current study will add to the weight of evidence suggesting agencies should consider such efforts. While the positive effects here were small, with more engaged participants those effects should be larger. Also, even small effects over large numbers of partici-

pants and repeated episodes of deliberation could help create a more informed and engaged public that grants agency decisions greater legitimacy. Perhaps as well, technologists should pay more attention to creating NLP applications that help agencies intelligently deal with their substantial information load from public comments, thus encouraging agencies to buy into these efforts.

Disappointingly, our NLP technologies had no significant effects on policy knowledge, policy attitudes, agency confidence, or perceived procedural justice. Also, with two of three indicators of improved citizenship the technology experimental condition had no significant effect. With respect to a third indicator, Citizen Identity, the technology condition showed a significant negative effect, though at only the .05 level. Perhaps this is a fluke consequence of looking at multiple indicators. There remains the concern, however, that something about the availability of the technology blocked the positive impact of deliberation on the centrality of citizenship to personal identity. Conceivably, NLP technologies that answer questions may give people less reason to consult others for information, thus reducing the stress on citizenship. Also, technologies that connect participants discussing similar topics might form subgroups of more like-minded individuals, again reducing opportunities to perceive a common citizenship. Either of these concerns would require additional research to establish firmly. Still, they suggest that technologists should be cautious about unintended consequences in complex social contexts. The MuLD technique more than reverses the negative impact of the technology, so it is possible to combine features to avoid negative effects.

Perhaps, the NLP technologies could have a positive effect with more engaged participants. In addition, improvements to the technology that would better help facilitate discussion may enhance positive effects. A full Discussion Facilitation Agent would include a variety of other technologies such as: summarization; argument maps; suggestions for additional lines of discus-

sion (for instance, summarizing the main points of a more successful discussion); feedback on discussion quality, possible miscommunication, and relevance; and requests that silent participants state what they think. Such functionality is within reach of existing NLP software, though such software would likely have to be tailored to and tested within the context of political discussion.

NLP technology could also be built to help agencies better cope with the large volume of public comments. Specifically, this includes an ability to highlight, list and potentially summarize all the arguments made in the deliberation. For instance, semantically similar comments could be clustered together to present an overview of all key points in the discussion. Along the same lines, software could classify the aspects of the discussion into the main areas covered during the deliberation. Finally, as previously indicated, technologists can apply polarity detection mechanisms commonly seen in opinion mining (Wiebe, Wilson, & Cardie, 2005) to determine the sides taken by participants in a deliberation. This would aid in automatically creating groups or units based either on similarity of opinion, or ensuring smaller groups reflect all opinions—as intended in MuLD.

Finally, the results here provide evidence for two positive effects of the MuLD technique. MuLD reverses the adverse effect of technologies on Citizen Identity. MuLD also significantly reduces right-wing authoritarianism, which has a number of negative effects on citizenship. MuLD did not result in more engaged participants being chosen to represent each group nor did it enhance policy knowledge, policy attitude change, or perceived legitimacy. These were the main effects expected, thus again suggesting that the low engagement of study participants may have insured minimal impacts of MuLD. The two significant effects observed may be "minimal manipulation effects." That is, simply by having participants undergo the MuLD technique, the researchers are suggesting that such arrangements and what they imply are normatively desirable. By directly

electing their own leaders, MuLD participants may have inferred that authorities are chosen and thus should not be blindly obeyed—mildly undermining authoritarianism. Also, the MuLD election mechanism stresses the role of the participant as part of a collective, perhaps thereby enhancing Citizen Identity.

The results here are preliminary in nature with respect to the implications of citizen e-deliberations as a tool of government. This study presents various dimensions and measures of evaluation and outcomes that may be fruitful for such work. The results suggestive of impacts of minimal manipulations indicate that technologists and deliberation designers must be careful regarding the implications that participants might draw from their designs. In general, the findings here are sufficiently encouraging to recommend continued work in this area, particularly work with real agency rulemakings.

ACKNOWLEDGMENT

REFERENCES

Barabas, J. (2004). How deliberation affects policy opinions. *The American Political Science Review*, *98*(4), 687–701. doi:10.1017/S0003055404041425

Benhabib, S. (1994). Deliberative rationality and models of democratic legitimacy. *Constellations (Oxford, England)*, *1*(1), 26–52. doi:10.1111/j.1467-8675.1994.tb00003.x

Bohman, J. (1996). *Public deliberation: Pluralism, complexity, and democracy*. Cambridge, MA: MIT Press.

Campbell, D. T., Stanley, J. C., & Gage, N. L. (1966). *Experimental and quasi-experimental designs for research*. Chicago, IL: R. McNally.

Chambers, S. (1996). *Reasonable democracy: Jurgen Habermas and the politics of discourse*. Ithaca, NY: Cornell University Press.

Dryzek, J. S. (2002). *Deliberative democracy and beyond: Liberals, critics, contestations. Oxford Political Theory*. Oxford, UK: Oxford University Press.

Endenburg, G. (1998). *Sociocracy: The organization of decision-making*. Delft, The Netherlands: Eburon.

Farrar, C., Fishkin, J. S., Green, D. P., List, C., Luskin, R. C., & Levy Paluck, E. (2010). Disaggregating deliberation's effects: An experiment within a deliberative poll. *British Journal of Political Science, First View*, 1-15.

Fishkin, J. S. (1997). *The voice of the people: Public opinion and democracy*. New Haven, CT: Yale University Press.

Gastil, J., Black, L. W., Deess, E. P., & Leighter, J. (2008). From group member to democratic citizen: How deliberating with fellow jurors reshapes civic attitudes. *Human Communication Research*, *34*(1), 137–169. doi:10.1111/j.1468-2958.2007.00316.x

Gutmann, A., & Thompson, D. (1996). *Democracy and disagreement*. Cambridge, MA: Harvard University Press.

Hardy, H., Shimizu, N., Strzalkowski, T., Liu, T., Zhang, X., & Wise, G. B. (2002). Cross-document summarization by concept classification. *SIGIR 2002: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* August 11-15, 2002, Tampere, Finland (pp. 121-128).

Luskin, R. C., Fishkin, J. S., & Jowell, R. (2002). Considered opinions: Deliberative polling in Britain. *British Journal of Political Science*, *32*(3), 455–488. doi:10.1017/S0007123402000194

Muhlberger, P. (2005). *Democratic deliberation and political identity: Enhancing citizenship*. International Society of Political Psychology 28th Annual Scientific Meeting. Toronto, Ontario.

Muhlberger, P. (2007). *Report to the Deliberative Democracy Consortium: Building a deliberation measurement toolbox*.

Muhlberger, P. (2011). (Manuscript submitted for publication). Stealth democracy: Authoritarianism and democratic deliberation. *Political Psychology*.

Muhlberger, P., Stromer-Galley, J., & Webb, N. (2011). *Public policy and obstacles to the virtual agora: Insights from the deliberative e-rulemaking project*. Information Polity.

Muhlberger, P., & Weber, L. M. (2006). Lessons from the Virtual Agora Project: The effects of agency, identity, information, and deliberation on political knowledge. *Journal of Public Deliberation*, *2*(1), 1–39.

Pivato, M. (2007). *Pyramidal democracy.* Retrieved from http://mpra.ub.uni-muenchen.de/3965/1/MPRA_paper_3965.pdf

Price, V., & Cappella, J. N. (2002). Online deliberation and its influence: The electronic dialogue project in campaign 2000. *IT & Society, 1*(1), 303–329.

Schweiger, D. M., Sandberg, W. R., & Ragan, J. W. (1986). Group approaches for improving strategic decision making: A comparative analysis of dialectical inquiry, devil's advocacy, and consensus. *Academy of Management Journal*, *29*(1), 51–71. doi:10.2307/255859

Small, S., & Strzalkowski, T. (2009). HITIQA: High-quality intelligence through interactive question answering. *Journal of Natural Language Engineering: Special Issue on Interactive Question Answering*, *15*(1), 31–54.

Stromer-Galley, J., & Muhlberger, P. (2009). Agreement and disagreement in group deliberation: Effects on deliberation satisfaction, future engagement, and decision legitimacy. *Political Communication*, *26*(2), 173–192. doi:10.1080/10584600902850775

Tyler, T. R. (2006). *Why people obey the law (illustrated edition.)*. Princeton University Press.

Webb, N. Hepple, M., & Wilks, Y. (2005). Dialogue act classification based on intra-utterance features. *Proceedings of the AAAI Workshop on Spoken Language Understanding*.

Webb, N., & Ferguson, M. (2010). Automatic extraction of cue phrases for cross-corpus dialogue act classification. *Proceedings of the 23rd International Conference on Computational Linguistics (COLING-2010)*. Beijing, China.

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Journal of Language Resources and Evaluation*, *39*(2-3), 165–210. doi:10.1007/s10579-005-7880-9

Zillig, L. P., Herian, M., Abdel-Monem, T., Hamm, J., & Tomkins, A. (2010). Public input for municipal policymaking: Engagement methods and their impact on trust and confidence. *ACM International Conference Proceeding Series: Proceedings of the 11th Annual International Digital Government Research Conference* (Vol. 292, pp. 41-50).