2006

Whither Deliberation? Mass e-Mail Campaigns and U.S. Regulatory Rulemaking

Stuart W. Shulman
University of Massachusetts - Amherst, stu@polsci.umass.edu

Whither Deliberation? Mass E-Mail Campaigns and U.S. Regulatory Rulemaking

Stuart W. Shulman

ABSTRACT. Mass e-mail campaigns are the organizational tool of choice for environmental activists seeking to inform and mobilize their constituencies. Some democratic theorists and reformers pin their hopes for more responsive and informed government policy on Internet-enhanced dialogue and debate. Electronic advocacy campaigns and action alerts are changing the nature and scope of public deliberation in contentious federal rulemaking. This paper examines the new digital landscape

Dr. Stuart W. Shulman is Assistant Professor with a joint appointment in School of Information Sciences and the Graduate School of Public and International Affairs at the University of Pittsburgh. He is Senior Research Associate in the University of Pittsburgh's Center for Social and Urban Research (UCSUR) and founder and director of UCSUR's Qualitative Data Analysis Program (QDAP). He has been the Principal Investigator and Project Director on related National Science Foundation-funded research projects focusing on electronic rulemaking, language technologies, digital citizenship, and service-learning efforts in the United States. He has been the organizer and chair for federal agency-level electronic rulemaking workshops held at the Council for Excellence in Government (2001), the National Defense University (2002), the National Science Foundation (2003), and The George Washington University (2004).

Address correspondence to: Dr. Stuart W. Shulman, 121 University Place, Suite 600, Pittsburgh, PA 15260 (E-mail: Shulman@pitt.edu).

This article was originally prepared as a talk at the 2004 Annual Meetings of the American Political Science Association in Chicago, IL, and adapted for the December 13-14, 2004, Institute for Public Administration of Canada "International Brainstorm: The Regulatory State of Tomorrow." This research was made possible with grants from the National Science Foundation (IIS-0429293 and SES-0322662). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect those of the National Science Foundation.

Journal of E-Government, Vol. 3(3) 2006
Available online at http://jeg.haworthpress.com

doi:10.1300/J399v03n03_03

of electronic rulemaking through a case study of the recent mercury rulemaking at the US Environmental Protection Agency. Results are presented from an effort to code a sample of 1,000 e-mails selected at random from a dataset of approximately half a million. Particular attention is paid to the unique additions made to form letters. doi:10.1300/J399v03n03_03

KEYWORDS. Internet, e-mail campaigns, public deliberation, electronic rulemaking, electronic advocacy, form letters

INTRODUCTION

Imagine that someone stood up at a professional meeting and said:

> If only information technology had been used to facilitate online dialogue and deliberation with the ENEMY, Al Queda, prior to 9/11, we might have resolved our differences. We might all have come to better understand the other sides' point of view. We might have collaborated to find a better solution than mass murder.

If such a thing were said, the audience would probably would howl and hiss in a manner not often seen at such meetings. This first argument is simply not credible. If instead they said:

> If only information technology could be better used to facilitate online dialogue and deliberation with the ENEMY, (say, industry, government, the environmentalists, Ralph Nader, or whomever) we might have resolved our differences. We might all have come to better understand the other sides' point of view. We might have collaborated to find a better solution than, say, allowing a suboptimal dose of mercury pollution into our nation's air that may ultimately harm or kill more people than died on September 11, 2001.

A few observers might endorse the second argument. Perhaps some would call it democracy's technological cutting edge (Berkman Center,

2005; Noveck, 2004a; Froomkin, 2004; Blumler & Coleman, 2001). That is, the potential in the United States for using information technology and Internet-enhanced participation in the "notice and comment" stage of the regulatory rulemaking process to make it widely distributed, reflexive, transparent, information rich, asynchronous, low-cost, less adversarial, and more meaningful (GAO, 2005; Bimber, 2003; Brandon & Carlitz, 2002; Johnson, 1998). Others more broadly hope for online democratic systems less encumbered by the exercise of power, intimidation, deception, single-mindedness, and other forms of even more commonplace speech-treachery (Parker, 2005; Beierle, 2003; Carlitz & Gunn, 2002; Coleman & Gøtze, 2001; Dahlberg, 2001).

It is only partly ridiculous to compare the dynamics of Jihadist zealotry with the behavior of sometimes visibly desperate actors in U.S. regulatory rulemaking. The passion to win, and for many the accumulated experiences of repeatedly losing, in an intensely adversarial and political environment has left significant numbers of citizen participants in a state of mind in which they are more likely to eviscerate or humiliate than deliberate. Online participation in U.S. rulemaking, at least for the foreseeable future, is no more likely to transform how the average citizen helps make regulatory decisions in the United States than it is to resolve the many issues underlying our current war on terror. While the potential for more meaningful forms of online deliberation is there (Shane, 2005; Stanley et al., 2004), for most participants in the heavy-traffic U.S. rulemakings, it is latent and undeveloped.

Optimists argue that online deliberation will indeed be transformative someday, but only when it is structured appropriately by the right people (Noveck, 2004b; OECD, 2003). I remain less sanguine, however, based on the state of the art in electronic rulemaking, current trends in e-advocacy, and because agreeing upon the "right people" is terribly difficult (Shulman, 2004, 2005). Furthermore, in the specific domain of mass participation in environmental rulemaking, it is quite possible that e-mail campaigns are another indicator of the "Death of Environmentalism," in which the mechanistic sustenance of the movement supersedes the type of thinking and action needed for the actual saving of the planet (Shellenberger & Nordhaus, 2005; Emery & Emery, 2005). At the very least, the domain of e-mail and cyberspace more generally is better known for flourishes of incivility rather than any widespread, technologically-driven transformation of passive citizens into thoughtful, deliberative, and engaged actors.

These cautious observations grow out of almost five years of work on what became two related but distinct National Science Foundation

(NSF)-funded projects. The first began with discovery of the National Organic Program rulemaking public comment data online (Shulman, 2003). In that instance, the United States Department of Agriculture (USDA) received approximately 277,000 comments on their initial proposed standard for the term "organic." Of those comments, more than 20,000 were submitted via a Web form set up ad hoc by mid-level USDA personnel for a 90-day comment period. Respondents using this online interface had the option to see comments that had already been submitted, no matter whether they came by the Web, fax, or postal mail, and to enter their own comments via a Web form.

In early 1998 it was a substantial innovation at USDA to allow online commenters anywhere in the world a chance to see the comments. The organic rule writers, managers, and technicians interviewed in 2000 and again in 2003 reported that this "open docket" design resulted in a significant number of comments on other peoples' comments, which was suggestive of nascent, large-scale, online dialogue and deliberation in federal rulemaking. Intrigued by this possibility, a group was formed to investigate whether this was part of a turn to deliberative democratic practices facilitated by the birth of e-rulemaking (Shulman et al., 2003).

In the first project, under the direction of David Schlosberg, a political theorist at Northern Arizona University, we are systematically looking for indicators of deliberation in a set of rulemakings in which the presence of an online open docket and a controversial environmental issue created at least the possibility for some degree of online deliberation amongst commenters with divergent points of view. In 2003 and 2004, a total of 10 undergraduates at two universities (Drake University and University of San Francisco) were professionally trained in two-day seminars to code a large, random sample of public comments from three rulemakings where the presence of an open docket created the possibility for deliberation.

Specifically, we were looking for signs in the text submitted that commenters had read other comments, or had otherwise demonstrated behavior that theorists of deliberative democracy call desirable. Therefore, we looked for signs of deliberation instead of preference aggregation, inclusion of difference, respect for a variety of positions, transformation of preferences, as well as expanding and authentic discourse. To date, there is very little evidence that the actual public comment text submitted meets any of our deliberative thresholds. Despite the presence of an online open docket, the technology itself fails to render commenters more deliberative.

These early findings from this textual analysis stand in contrast with the self-reported deliberative behavior of a majority of respondents in a large-scale (1,553 respondents) telephone survey of commenters on these same rules (Schlosberg et al., 2005). It may suffice to say for now that the specific barriers to better online deliberation, at least in U.S. regulatory rulemaking, are less technical than social, political, legal, and architectural in nature.

In the other project, with a pair of computer science colleagues, we are developing and testing the efficacy of new information retrieval tools tailored to the rulemaking environment (Shulman et al., 2004, 2005a). As part of this project, we convened a series of workshops, interviews, and focus groups annually for four years with governmental and non-governmental stakeholders involved in regulatory rulemaking. At the most recent of these workshops, we demonstrated the state of the art in our research and development of algorithms that can, among other things, quickly and efficiently identify duplicate electronic form letters and unique text added to near-duplicate comments, and then we ask how the availability of these and other advanced language processing tools might impact the rulemaking process (Shulman, 2004).

The reason we do this is that agency personnel have repeatedly told us that the emergence of first generation electronic rulemaking has had the singular effect of increasing the flood of duplicative, often insubstantial, mass mailing campaigns (Shulman et al., 2005b). These campaigns exist for many legitimate reasons, politically and organizationally speaking. However, they do little to move administrative rulemaking toward the ideal of enlightened online deliberation and they do much to try and swing the pendulum away from administrative expertise and toward plebiscitary, direct democracy via electronic preference aggregation. This terrifies administrative law scholars and practitioners in the U.S. civil service as the "overall trend has been away from the expertise model and toward the politics model," as the comments submitted increase by orders of magnitude (Lubbers, 2003: 149-150). It also fails to move the process to a higher deliberative plane.

So, where is this heading? Interest groups contract with a thriving e-advocacy sub-sector of the Washington, DC economy and routinely set up clusters of Web action centers. The prevailing system generates voluminous quantities of mass e-mail comments. One firm (Capitol Advantage, see *http://capitoladvantage.com*) promoted itself in 2004 with a counter on their home page claiming credit for more than 16 million constituent messages that year. This brain-numbing quantity of duplicative e-mails leaves practitioners in federal agencies (or more commonly

private contractors) sorting e-mails with mounting congressional and executive demands for efficiency and effectiveness and a highly charged, adversarial political and ideological backdrop. Considering that billions of dollars and the nature of life on earth are at stake, the fact is that the deliberative democratic future of regulatory rulemaking looks bleak.

These machinations occur, in part, for reasons that have little to do with improving the final rule. While a Capitol Advantage service like "CapWiz" offers customers the technical ability to efficiently deliver a large number of messages that are "on time, on target, and on point," there is more than simply effective communication at stake. Modern data mining and outreach techniques mean that successful campaigns can significantly increase membership lists and donations. MoveOn.org is the best illustration of this phenomenon, but hardly the only one. A wide rage of interest groups seek free media and other forms of Internet-driven publicity (e.g., coverage in blogs) when they initiate a public comment campaign. Large public comment outpourings also may work as delay and litigation tactics, as well as congressional wake-up calls to revisit an issue. These are only a few of the many reasons for the proliferation of duplicative, insubstantial electronic postcards that federal officials speak dismissively of as insubstantial "awareness" campaigns.

Some groups believe that overwhelming an agency like the Forest Service with form letters will result in a favorable outcome. Devotees to one or the other side in the battle over the Roadless Conservation Area rulemaking might, at this late date, wonder whether all the spilled ink and clicked send buttons have actually been efficacious in the face of mountains of litigation that keep the rule in the courts. While these campaigns are presumed to be largely ineffective because they generate little new information, in some instances (e.g., the EPA's Advanced Notice of Proposed Rulemaking on the Definition of U.S. Waters, or the USDA's organic rulemaking, where about 100,000 unique comments carried weight with officials and influenced the final rule), it does at least appear to contribute to an outcome favorable to the mass mailers. The organizational incentives combined with occasional claims of victory suggest the practice is likely to dominate the near future of electronic rulemaking.

CURRENT RESEARCH ACTIVITY

On January 20, 2004, the Environmental Protection Agency (EPA) published a Notice of Proposed Rulemaking titled "Proposed National

Standards for Hazardous Air Pollutants."[1] In response, the agency received approximately 490,000 public comments.[2] An EPA press release dated December 15, 2003, had then Administrator Mike Leavitt predicting the proposed actions would be "the largest air pollution reductions of any kind not specifically mandated by Congress."[3] Precisely what Congress mandated remains a matter of considerable legal debate.

Like many significant federal regulatory actions, the EPA's Clean Air Mercury Rule resulted from a drawn out mix of congressional, administrative, and legal proceedings stretching back to the Clean Air Act (CAA) Amendments of 1990. A lawsuit by the Natural Resources Defense Council filed in 1992 and another by the Sierra Club in 1994 were settled later in 1994 and ultimately resulted in a "Mercury Study Report to Congress" (RTC), which was released in December, 1997.[4] In the EPA's RTC, it was estimated that "about three times as much mercury is being added to the global reservoir from U.S. sources as is being deposited from it."

On December 14, 2000, one day after Al Gore conceded the 2000 election, outgoing EPA Administrator Carol Browner announced a finding that it was "appropriate and necessary" to regulate coal- and oil-fired electric utilities under Section 112 of the CAA. Section 112 requires that industry limit hazardous air pollutants using the maximum achievable control technology. On March 15, 2005, acting EPA Administrator Steve Johnson signed the final "Clean Air Mercury Rule," noting the United States was "the first nation in the world to address this remaining source of mercury pollution."[5] The new rule set out a "cap and trade" system, modeled on earlier success with reductions in acid rain-causing sulfur dioxide and nitrogen oxides.[6] According to the EPA, the program was designed to reduce mercury emissions from coal from coal-fired power plants by 70% in the year 2018.[7] Johnson's press release argued that while airborne mercury was indeed a matter of great concern, particularly for women of child-bearing ages, the United States as a whole is responsible for only 3% of the global human-caused emissions, and the utilities themselves produce only 1% of global human-caused mercury emissions.

And yet the health and environmental impacts of mercury in the United States are substantial. In 2003, the EPA reported "48 states, the District of Columbia and American Samoa issued 3,094 fish advisories, 280 more than the previous year. With these additions, 35% of the total lake acres and 24% of the river miles in the nation are now under advisory."[8]

On May 18, 2005, in what might be termed a (quite clearly Democratic) state coalition legal strategy, Attorneys General of nine states and an Environmental Protection Secretary from a tenth state, joined with New Jersey Attorney General Peter Harvey filing suit in the DC Circuit Court to block the Bush administration rule.[9] Harvey's New Jersey Department of Justice news release was widely quoted in the press with the Attorney General asserting the new rules were "deeply flawed and contrary both to science and the law."[10] In the announcement of their joint lawsuit, the coalition filing suit noted the rule will "delay meaningful emission reductions for many years and perpetuate hot spots of local mercury deposition, posing a grave threat to the health of children." In addition, a dozen national and regional environmental and conservation organizations are collaborating on a total of four separate lawsuits in federal courts, arguing mercury hot spots will disproportionately impact individual communities.[11]

Advocates of more stringent controls in the press were on guard as a result of a tendentious rulemaking process marked by irregularities. For example, a report by the EPA's Inspector General found senior agency officials manipulated data and scientific studies to reach a predetermined pro-industry outcome.[12] A General Accountability Office report called the EPA's economic analysis "of limited use," listing among its four major shortcomings the failure to "estimate the economic benefits directly related to decreased mercury emissions."[13]

After the final rule promulgation, editorial pages across the United States expressed rueful dismay or else fully lambasted the Bush Administration on substantive and procedural grounds. Editorial writers at *The Boston Globe*, having learned that an EPA-commissioned Harvard study estimating health benefits of an alternative proposal at 100 times the level used by the EPA was "stripped from public documents," opined that the agency "cooked the books to arrive at its weak rule."[14] At *The Baltimore Sun*, the editorial page put it even more baldly: "President Bush doesn't like regulations, doesn't like treaties and doesn't like scientific research that contradicts his views. The result could be a record crop of brain-damaged babies."[15] The *Portland Press Herald* reminded readers that a *Los Angeles Times* investigation revealed the proposed rule used verbatim language from industry sources, concluding the "way these rules were developed is nothing short of laughable."[16] The *Milwaukee Journal Sentinel* wrote the rules "could well be a tragedy," whereas the *Pittsburgh Post-Gazette* decried the "porous plan" as a "new but not unexpected approach" likely to leave regions

like southwestern Pennsylvania saddled with the ill-effects of toxic mercury hot spots.[17]

Data and Methodology

In August 2004, the EPA provided the author with 536,975 text files presumed at that time to be the entire population of public comments submitted via e-mail in response to the Notice of Proposed Rulemaking.[18] A random sample of 1,000 documents was selected for systematic qualitative data analysis. Five graduate student coders and a project manager were trained by an outside consultant and the author to complete the coding at the University of Pittsburgh during the fall of 2004.

At the beginning of each file, e-mail header text was present that was not going to be coded. A small shell script was written by the graduate student project manager and run in UNIX to remove all header text lines. Once all header text lines were removed, a small macro was recorded to remove all unnecessary hard returns from the text files, rendering them more amenable to analysis using ATLAS.ti, a qualitative data analysis software application capable of handling large scale projects with multiple coders.[19]

Document Division

The goal of the document division was to divide the 1,000 documents among the five coders so that the file number sequences were mixed in a way that decreased potential problems of similarity due to sequential order. The documents were divided in a way that they overlapped 50%, ensuring that the overlapped document for each coder was coded by two other coders (see Table 1). Sixty documents were used for each round of coding. Fifteen of each 60 documents were coded by a second coder and another 15 were coded by a third student allowing a systematic test of the inter-rater reliability (see section 2.4). Hence, a total of 320 documents were coded twice while 680 were coded only once.

Using an off-the-shelf qualitative research software package, ATLAS.ti, the coders completed the first round of coding extremely quickly. This was due to a large percentage of exact duplicate form letters (discussed further in section 3.2) and the brevity of the unique additions to form letters. Some coders completed the first set of 60 documents within an hour and a half. Students were trained to identify passages of text that corresponded with one or more predetermined codes (applied at the sentence, multi-sentence, or paragraph level) and families (applied at the

TABLE 1. Sample of Document Division Strategy for Inter-Rater Reliability Tests

Coder	Document IDs	Overlap	Coded By	Unique Documents
Pitt1	01-60	01-15, 46-60	Pitt5, Pitt2	16-45
Pitt2	46-105	46-60, 91-105	Pitt1, Pitt3	61-90
Pitt3	91-150	91-105, 136-150	Pitt2, Pitt4	106-135
Pitt4	136-195	136-150, 181-195	Pitt3, Pitt5	151-180
Pitt5	01-15, 181-225	01-15, 181-195	Pitt1, Pitt4	196-225

document level). Appendix A presents the complete code and families list with definitions.

Once the first round of coding was complete, the project manager conducted an analysis to verify whether codes and families were added appropriately. Examining random documents from each user, it was evident that all of the coders lacked a clear understanding of at least some of the code definitions, their proper application, and the best routine for ensuring that every applicable code and family designation was assigned. Based on an analysis of the error, a supplementary training document was created consisting of screen shots and textual analysis of where codes were not used or inappropriately assigned. The coders reviewed the document to see where they and their peers diverged in the application of the coding strategy. After each coder completed the review, they received a second round of 60 documents. Three of the five coders completed four rounds where the other two coders completed five rounds.

Upon completion of enough rounds of coding to finish the sample of 1,000 documents, two data sets were created. The first contained the 680 documents that had been coded by a single coder, the second contained 320 pairs of identical documents that had been coded by two coders. A second review stage of the second set of overlapping documents revealed persistent inconsistencies in the coding. The coders had the most difficulty with the codes "Comment on a Comment," "Information in Docket," "Stakeholder," "Stock Phrase," and "Strength High." The coders also failed to assign the appropriate document-level families. After a thorough review, a third re-training session was conducted to clear up lingering confusion about codes and families.

Following the re-training, the coders returned to their assigned documents and reviewed their coding, independently correcting their mistakes,

and adding codes and families where they were missing. Each coder's set of documents was subsequently reviewed by at least two other coders for omissions before a final master data set was created.

Inter-Rater Reliability

Tools for reviewing and coding text, such as those found in a qualitative data analysis package like ATLAS.ti, carry with them no particular attributes of reliability. Using computers does not make the coding itself reliable. It is better to speak of the reliability of the results from a particular set of observations, rather than the tool itself, the coding system, or its developers. The benefits of using ATLAS.ti are nonetheless significant. It allows for scalable, replicable, transparent, and multi-coder passes over substantial quantities of text, with significant opportunities for collaborative learning via the pre-test methods discussed earlier, as well as the possibility of systematic tests of rater reliability once the disaggregated coding has been merged together.

Empirical researchers once assumed that "direct observation of behavior was by definition bias-free and valid" (Suen & Ary, 1989, 100). Current research methods assume bias and error and seek evidentiary-based claims about the reliability of observational data. The goal is to show consistency in results over numerous repetitions or, better still, through independent observations.

Statistical techniques have been developed to assess the reliability of such observations. The psychometric and statistical literature refers to inter-rater and inter-observer measures interchangeably. According to Suen and Ary (1989, 103), agreement is a measure of "the extent to which two or more observers agree on the occurrences and non-occurrences of a behavior." In this study, the behavior being observed consists of the use of one or more rhetorical strategies for influencing rule writers at the EPA.

A number of different techniques have been developed, including the percentage agreement index, the occurrence and nonoccurrence agreement index, and the kappa coefficient (von Eye & Mun, 2005). Studies in the area of child development and psychology in the late 1970s showed that the percentage agreement statistic, which was, at the time, an easy to calculate yet highly controversial statistic, was also the most popular technique for reporting inter-rater reliability (Suen & Ary, 1989). Critics raised concerns about spurious inflation due to chance,

while proponents defended the technique as "parsimonious both in concept and in computational procedures" (108). It is computed as follows:

$$\text{Percentage Agreement } (p) = \frac{\text{Number of Agreements } (n)}{n + \text{Number of Disagreements}}$$

Over recent years, "coded corpora have acquired an increasing importance in almost every aspect of human language technology," (Di Eugenio, 2000: 441) as well as in a range of disciplinary and interdisciplinary research efforts. The ongoing collaboration with the computer science research community called for a more precise reliability metric, as well as a tool for generating it easily and at various points in the project. Cohen's (1960) kappa coefficient is widely considered a better standard measure of the degree of agreement existing beyond chance alone across a wide range of annotation efforts. The general equation for kappa is:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where p_o is the observed proportion agreement and p_e is the expected proportion agreement by chance (Suen & Ary, 1989; Sim & Wright, 2005).

For our work on public comment discourse analysis, a useful annotation analysis tool was developed and deployed on the Web by a graduate student computer science researcher and a programmer at The University of Southern California, Information Sciences Institute (USC-ISI).[20] The tool is for reporting the kappa coefficient and agreement index scores between two sets of annotations on a single set of documents. The kappa coefficients presented in this paper are based on only the coded output of an ATLAS.ti project. These scores do not reflect computations incorporating the full corpus of text, including that which received no codes, which is an alternative method for computing kappa we are exploring. The terms that follow explain the output reported in Tables 2 and 3:

- *Exact Match:* Indicates the raw number of times both coders identified identical text segment spans (with not a single character or white space of variation) and assigned the same code. At the time of the original coding, the kappa tool had not yet been built and the need for high precision in text spans was less well understood.

- *Overlap:* Indicates the raw number of times when matching codes were applied to text segment spans that overlap but do not have exact boundary matches.
- *Kappa:* An index with a range −1.0-1.0 that measures agreement when both coders do or do not apply a code. A positive kappa indicates that observers agree more than they would by chance. A score of 0.8 or higher is considered a high level of agreement, whereas above 0.6 is considered substantial agreement (Landis & Koch, 1977; Krippendorf, 1980). Any such heuristic, however, may "oversimplify the interpretation of the values of kappa" (Suen & Ary, 1989: 113). Some statisticians argue you need to consider the cost of reaching wrong conclusions. Personal experience with this tool indicates it is a dubious indicator for specific codes when the raw counts are only a small portion of the coded text.
- *F-measure:* An alternate index with a range −1.0-1.0 that is proving more useful for reporting inter-rater reliability. When comparing two coder's annotations, the F-measure regards one set of annotations as the correct answer and the other as the coding system output. The equation, therefore assumes P is equal to (match)/(total number for coder 1) and R is equal to (match)/(total number for coder 2), allowing us to calculate an F-measure as $(P * R * 2)/(P + R)$.

TABLE 2. Aggregated Kappa Coefficient and F-Measures for Major Codes

Code	UCSUR 1	UCSUR 2	Exact Match	Overlap	Kappa (κ)	κ (w/ overlap)	F-Measure	F-Measure (w/overlap)
Disappointment	77	78	26	15	0.66	0.75	0.34	0.53
Economic	62	88	20	24	0.63	0.78	0.27	0.59
External authority	35	68	9	10	0.72	0.78	0.17	0.37
Legal	33	53	5	9	0.75	0.80	0.12	0.33
Proposal	115	185	27	37	0.18	0.39	0.18	0.43
Public health and safety	151	216	65	60	0.21	0.57	0.35	0.68
Science and technology	92	144	23	25	0.37	0.51	0.19	0.41
Social values	125	153	53	49	0.43	0.73	0.38	0.73
Stakeholder	19	28	5	13	0.88	0.96	0.21	0.77
Unique text ("Form+")	72	68	49	13	0.86	0.95	0.70	0.89
Total	**781**	**1081**	**282**	**255**	**0.57**	**0.72**	**0.30**	**0.58**

TABLE 3. Kappa Coefficients and F-Measures for Sub-Coding of "Unique Text from Form Letters"

Code	UCSUR 8	UCSUR 9	Exact Match	Overlap	Kappa (κ)	κ (w/overlap)	F-Measure	F-Measure (w/overlap)
Agency mission	51	42	38	0	0.82	0.82	0.82	0.82
Anecdotes	45	24	23	0	0.76	0.76	0.67	0.67
Catering to business	55	45	38	2	0.75	0.79	0.76	0.80
Children's health	86	81	75	5	0.82	0.93	0.90	0.96
Disbelief	35	20	16	1	0.76	0.78	0.58	0.62
Economic	14	36	9	2	0.67	0.71	0.36	0.44
Higher values	33	10	5	0	0.66	0.66	0.23	0.23
Insults	27	32	21	0	0.82	0.82	0.71	0.71
Public interest/ health	121	59	50	7	0.18	0.29	0.56	0.63
Sarcasm	13	12	5	1	0.85	0.85	0.40	0.48
Science	5	11	5	0	0.94	0.94	0.62	0.62
Shame	44	8	8	0	0.63	0.63	0.31	0.31
Substantive claim	24	0	0	0	0.75	0.75	0.00	0.00
Toxics	39	25	20	4	0.75	0.83	0.62	0.75
Total	**595**	**405**	**313**	**22**	**0.73**	**0.76**	**0.63**	**0.67**

In this study, 320 of the original sample of 1,000 e-mails had two sets of annotations. Under the document division strategy outlined in Table 1, a unique set of two of the five Pitt coders were responsible for each pair of double-coded documents. To generate the overall kappa coefficient and F-measures, the annotation "Author" ($Pitt_{1\text{-}5}$) was changed to UCSUR 1 for one set of 320 completed documents and to UCSUR 2 for the other set. Hence, Table 2 aggregates all measures of inter-rater reliability between the five coders over the full set of 320 double-coded documents.

Less frequently applied codes were removed from the project prior to generating the ATLAS.ti report ("All Quotations") on which Tables 2 and 3 are based. The rationale behind this decision is that, in those instances, the number of observations of so-called minor codes was so

few that the "agreement" of non-observations resulted in extremely high kappa coefficients. To produce a table with summary totals undistorted by inflated scores on the minor codes, these codes were deleted from a duplicate version of the project and then the report was generated for analysis using the USC-ISI kappa tool.

For the purposes of this study, two observations stand out. First, for a number of the codes that were applied frequently enough to produce valid inter-rater reliability scores (including overlaps), there is significant agreement. The iterative process of revisiting the coding definitions and examples of mismatch along the way resulted in solid coding of concepts such as "Social Values," "Stakeholder," and "Public Health and Safety." Second, and more relevant to the focus of this paper, the inter-rater reliability measures for the code "Unique Text in a Form Letter" were 0.95 (kappa including overlap) and 0.89 (F-measure including overlap). Therefore, in our study of 1,000 e-mails, the Pitt graduate student coders were fairly reliable when it came to isolating unique additions to form letters.[21]

INITIAL FINDINGS

At this early stage in the research (and in the epoch of mass e-mail campaigns) there are few indications that online deliberation is enhanced within the current e-government configuration in the United States. The mass e-mail campaign in particular appears to be an odd and possibly counter-productive tribute to twentieth century notions of one-directional, non-deliberative, and un-reflexive nose counting. Although the medium could be used to promote better dialogue, debate, deliberation, and public understanding, it falls short of the loftier ideals held out by hopeful political theorists.

E-Mail Is Not a Good Medium for Deliberative Acts, Such as Comments on Other Comments Already Submitted, or Respect for Different Points of View

If one is looking for indications of online deliberation, it simply does not happen via e-mail, at least not in this particular rulemaking. The code designed to capture deliberative behavior, "Comment on a Comment," was applied only once. The one instance in 1,000 documents of this code was: "You already have the boilerplate info from other letters, which I won't repeat here. Be assured, I DO SHARE THOSE

CONCERNS." This commenter went on to write a comment that included several of the most important stock phrases from form letters and nothing original.

Looking back, it should have been more obvious at the outset that e-mail clients and Web advocacy services are not set up to promote reflection on, or responses to, comments from people with whom the commenter disagrees. The architecture of e-mail is structurally ill-suited for deliberation about the merits of a proposed rule. When a user of a Web advocacy form is constructing their unique addition to a form letter (which is delivered as an e-mail), the response is not to the actual proposed rule published in the *Federal Register* nor the reasoned claims of the other side. Rather, it is almost exclusively to the appeals and imagery of the advocacy campaign organizers.

Perhaps the only real possibility for finding any significant quantity of actual online deliberative behavior might be a sample drawn only from people who commented from inside an actual web-based system (such as EPA's E-Docket), or some other system that fits the architectural definition of an open docket. The net effect, at least in the mercury case, seems to be the rendering of e-mail as a de facto medium for engaging in perfunctory assaults on the logic, science, competence, values, and virtues of the agency and the administration overall.

The Ability to Amend Form Letters Generated via Web Advocacy Campaigns Results in Very Few Substantial Additions to Pre-Formulated Awareness Campaign Text

Each of the documents was assigned to one or more families, such as exact duplicate, near duplicate (which are modified form letters), suspected duplicate, or off-topic (see Appendix B). A total of 67 documents were determined to be off-topic. Of these 67 off-topic files, most were comments on a different rule handled by the EPA's Office of Air and Radiation and a small number were commercial spam. After the completion of the document and coding merge, suspected duplicates were reviewed and assigned a proper family status by the author and all of the off-topic documents were permanently removed from the data set.

Among the 933 files that remained, 740 files (79.3%) were coded as exact duplicates generated by electronic advocacy campaigns. The only variations in these files were the submission dates and the commenters' personal identification data. Another 182 files (19.5%) were determined to be modified form letters, in which the commenter had added at least one sentence to a body of text that was otherwise duplicative.

Hence, according to the coding, a total of 922 of 933 e-mails in the sample (98.8%) were generated by electronic form letter campaigns.

The average length of the unique additions to form letters was approximately 38 words, or about the length of one very short paragraph. The following empirical section of this paper focuses on just those 182 unique additions to form letters that have come to be known as the "+" in "form +" letters. Having had the coders identify 182 "form +" letters in our sample of 1,000 e-mails, the complete "+" content was then subcoded by two different, independent coders using thematic categories created by the author after his review of the 182 instances of unique text. Several inter-related themes and very few substantive comments, by agency standards, were identified.

The quantitative totals in Table 4 report the instances of quotations assigned to a code by one or both of the coders engaged in the subcoding of this text. The author reviewed and verified every instance of the form + sub-coding; a number of quotations were "unlinked" when it was determined a code had been applied in error or as a result of a misjudgment. Hence, the quantitative data presented in Table 4 resulted from the author adjudicating disputes between the two independent subcoders (UCSUR 8 & 9).

TABLE 4. Independent and Final Adjudicated Counts

Code Name	UCSUR 8	UCSUR 9	Final
Agency mission	51	42	47
Anecdotes	45	24	36
Catering to business	55	45	51
Children's health	86	81	82
Disbelief	35	20	34
Economic	14	36	38
Higher values	33	10	32
Insults	27	32	30
Public interest/health	121	59	104
Sarcasm	13	12	18
Science	5	11	4
Shame	44	8	22
Substantive claim	24	0	8
Toxics	39	25	35
Total	**592**	**405**	**541**

Mass E-Mail Campaigns May Do More Harm than Good if They Make It Harder to Find the Useful Comments or Lower the Estimation of the Public Role in the Minds of Regulators

Even if there are a few more substantive (though not deliberative) comments in the mix, (and based on the mercury sample, perhaps less than 5% of every addition to a form letter may be substantive), the more the volume increases the more likely that anything good submitted will be lost under the current sorting regime, which consists of hiring for profit contractors and manually identifying "unique comments" by eye while looking at the printed version. Based on this preliminary exploration of the mercury sample, for every brief but substantive comment tacked on to a form letter e-mail, a reviewer will also have to read at least 90-95 pithy, pleading, condescending, name calling, or otherwise useless comments. It may therefore do more harm than good when hastily typed. Unreflective tirades are the bulk of the comments and they drown out the people whose carefully drawn comments might actually make a difference.

CONCLUSION

After a review of the data, it is difficult to be anything but cynical regarding the viability of finding deliberation in the current age of e-rulemaking. The rub is this: Whether for good or bad, the current system means that old-fashioned rules of thumb, like the 20-page rule, or the letterhead rule, are effective filters for officials who say they know in advance what they will need to read and what it will say. While the occasional lone voice speaks to us about the thrill of finding a gem in amongst a large number of duplicative thoughts, the agencies mostly farm that analytical work out to contractors and focus on what conventionally is known to matter. No doubt the best intentions of emotive, pleadings citizens will continue to result in floods of redundant comments; e-mail is a boon for generating those.

We in the e-rulemaking research group do not all reach the same conclusions about what is observable so far, neither in the data nor the responses of various actors involved. We are not without some shared hope, however, that innovation both at the federal level and in the NGO and e-advocacy sectors will eventually result in more meaningful online deliberation in controversial rulemakings. One can imagine deployment of more creative uses of IT by the groups to engage their members in

innovative, IT-enhanced efforts to distill the wisdom of the collective. For example, interest groups could retain their ability to mine the participants for data while getting 10, or 20, or even 50,000 people broken up into small groups that brainstorm, deliberate, and distill, then the groups aggregate into larger clusters of groups, then clusters of clusters, who all along can visualize via the Web, the best ideas, examples, and stories as they rise to the top. With a highly interactive goal in mind, you can imagine all sorts of Meetup.com style engagement add-ons and other innovative tools, like those developed by the "PICOLA" project at Carnegie Mellon University[22] and Beth Noveck's "Cairns" project at the New York Law School.[23] There are a number of possibilities about how this might emerge over time, yet to date, they remain largely the dreams of theorists who reside outside the beltway and whose work is at least one step removed from the actual rulemaking battleground.

NOTES

1. See *Federal Register* Vol. 69, No. 20 pp. 4652-4752 (available at: *http://snipurl.com/dab9*).
2. The EPA reports it "initially received over 680,000 letters, e-mails and postcards, including almost 5,000 unique messages, commenting on the Proposed Clean Air Mercury Rule and the Supplemental Clean Air Mercury Rule by the time the public comment period ended on June 29, 2004. As of February 2005, EPA E-Docket shows an actual count of more than 490,000 public comments and close to 4,500 unique comments received. The initial count of 680,000 and 5,000 included duplicate and triplicate e-mails and comments related to other rules." See: *http://snipurl.com/dabd*
3. See: *http://snipurl.com/f13z*
4. See: *http://www.epa.gov/mercury/report.htm*
5. See: *http://snipurl.com/f146*
6. See: *http://www.epa.gov/airmarkets/capandtrade/index.html*
7. The Natural Resources Defense Council disputes the claim, insisting the EPA's new rule will attain only a 50% reduction by 2020. See: *http://snipurl.com/fuir*. Under a cap and trade system, the total amount of mercury emissions declines nationally, however plants can either innovate to reduce their emissions or purchase excess credits in a quasi-market mechanism.
8. See: *http://snipurl.com/f1l6*
9. See: *http://snipurl.com/fuis*. The other states were: California, Connecticut, Maine, Massachusetts, New Hampshire, New Mexico, New York, Pennsylvania, Vermont, and Wisconsin.
10. Michelle O'Donnell, "11 States Challenge Break on Mercury for Coal Power Plants," *New York Times* (May 19, 2005).
11. See the joint environmental interest group press release at: *http://snipurl.com/fuiu*.
12. Shankar Vedantam, "EPA Inspector Finds Mercury Proposal Tainted," *Washington Post* (February 4, 2005), A4. Also see: *http://snipurl.com/fuiv*

13. The GAO report is online at: *http://www.gao.gov/new.items/d05252.pdf.*
14. Editorial, "Retrograde on Mercury," *Boston Globe* (April 1, 2005), A1.
15. Editorial, "Thumb on the Scale," *The Baltimore Sun* (March 25, 2005), 14A.
16. Editorial, "EPA's Mercury Rules are Just One Symptom of a Greater Problem," *Portland Press Herald* (March 30, 2005), A9.
17. Editorial, "Mercury Rule Too Weak," *Milwaukee Journal Sentinel* (March 28, 2005). Editorial, "Toxic Decision," *Pittsburgh Post-Gazette* (March 17, 2005), A14.
18. The full population of mercury e-mail is available for research purposes along with 11 other public comment text data sets at the e-rulemaking Testbed hosted by Carnegie Mellon University at: *http://hartford.lti.cs.cmu.edu/eRulemaking/Data.html*
19. In this and other data and project management tasks, as well as the writing of the first draft of sections 2.2 and 2.3, the work of Graduate Student Researcher Ahmed Shaffi was indispensable.
20. Many thanks to ISI's Namhee Kwon, a student of Dr. Eduard Hovy, who built the tool, and to Lei Ding, a staff programmer, who made it available on the Web at: *http://www.isi.edu/~nkwon/eRule/kappa.html.* This unsupported, experimental tool is available for any ATLAS.ti user who would like to review or report these inter-rater agreement statistics. The output here is based on uploading the "All Quotations" file from the ATLAS.ti quotation manager as plain text after the work of multiple coders had been merged together.
21. A number of interesting observations will follow in future papers about the task of producing more reliable coding for salient but diffuse concepts such as "Science and Technology," "External Authority," "Legal," and "Disappointment." Other ATLAS.ti output analysis tools currently under construction by USC-ISI personnel will allow better and faster training that will improve the overlap, increase the number of exact matches, and cut down single instance annotations that indicate a second coder missed an observation. In separate ongoing studies completed after this round of coding, significantly higher inter-rater reliability scores are being reported as a result of narrowing the number of codes in use and increasing the precision of the rule set governing the coding process.
22. See: *http://communityconnections.heinz.cmu.edu/picola/index.html*
23. See: *http://www.nyls.edu/pages/2150.asp*

REFERENCES

Beierle, Thomas C. 2004. "Digital Deliberation: Engaging the Public Through Online Policy Dialogues," in Peter Shane (ed.) *Democracy Online: The Prospects for Political Renewal Through the Internet* (New York: Routledge), 155-166.

Berkman Center for Internet & Society. 2005. "Online Deliberative Discourse Project," *http://cyber.law.harvard.edu/projects/deliberation/* [Last accessed February 28, 2005].

Bimber, Bruce. 2003. *Information and American Democracy: Technology in the Evolution of Political Power* (New York: Cambridge University Press).

Blumler, Jay G. and Stephen Coleman. 2001. "Realising Democracy Online: A Civic Commons in Cyberspace," IPPR/Citizens Online Research Publication No. 2.

Brandon, Barbara H. and Robert D. Carlitz. 2002. "Online Rulemaking and Other Tools for Strengthening Civic Infrastructure," *Administrative Law Review* 54(4), 1421-1478.

Carlitz, Robert D. and Rosemary W. Gunn. 2002. "Online Rulemaking: A Step Toward E-Governance," *Government Information Quarterly* 19, 389-405.

Coglianese, Cary. 2004. "E-Rulemaking: Information Technology and Regulatory Policy," Regulatory Policy Program Report No. RPP-05.

Coglianese, Cary. 2003. "The Internet and Public Participation in Rulemaking," Paper prepared for conference on Democracy in the Digital Age, Yale Law School, April 2003.

Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement* 20, 37-46.

Coleman, Stephen and John Gøtze. 2001. *Bowling Together: Online Public Engagement in Policy Deliberation http://www.bowlingtogether.net/* [Last accessed January 31, 2003].

Dahlbergh, Lincoln. 2001. "Computer-Mediated Communication and the Public Sphere: A Critical Analysis," *Journal of Computer-Mediated Communication* 7(1) *http://www.ascusc.org/jcmc/vol7/issue1/dahlberg.html* [Last accessed March 1, 2005].

Di Eugenio, Barbara. "On the Usage of Kappa to Evaluate Agreement on Coding Tasks," in *LREC2000, Proceedings of the Second International Conference on Language Resources and Evaluation, Athens, Greece*, 441-444.

Emery, Fred and Andrew Emery. 2005. "A Modest Proposal: Improve E-Rulemaking by Improving Comments," *Administrative and Regulatory Law News* 31(1), 8-9.

Froomkin, A. Michael. 2004 "Technologies for Democracy," in Peter Shane (ed.) *Democracy Online: The Prospects for Political Renewal Through the Internet* (New York: Routledge), 3-20.

Government Accountability Office. 2005. "Electronic Rulemaking: Progress Made in Developing a Centralized E-Rulemaking System," GAO-05-777.

Johnson, Steven M. 1998. "The Internet Changes Everything: Revolutionizing Public Participation and Access to Government Information Through the Internet," *Administrative Law Review* 50(2), 277-37.

Kerwin, Cornelius M. 1999. *Rulemaking: How Government Agencies Write Law and Make Policy* 2nd ed. (Washington, DC: CQ Press).

Krippendorf, Klaus. 1980. *Content Analysis: An Introduction to Its Methodology* (Beverly Hill, CA: Sage Publications).

Landis, J. Richard and Gary G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data," *Biometrics* 33, 159-174.

Noveck, Beth S. 2004a. "The Electronic Revolution in Rulemaking," which can be downloaded at *http://snipurl.com/88jf* [Last accessed February 28, 2005]

Noveck, Beth S. 2004b. "Unchat: Democratic Solution for a Wired World," in Peter Shane (ed.) *Democracy Online: The Prospects for Political Renewal Through the Internet* (New York: Routledge), 21-34.

Organization for Economic Co-operation and Development (OECD). 2003. *Promise and Problems of E-Democracy: Challenges of Online Citizen Engagement* (Paris: OECD Publications).

Parker, Richard. 2005. "21st Century Rulemaking: Executive Summary," presented at the American Bar Association's Administrative Law Conference.

Schlosberg, David, Stephen Zavestoski, and Stuart Shulman. 2005. "'To Submit a Form or Not to Submit a Form, That is the (Real) Question': Deliberation and Mass Participation in U.S. Regulatory Rulemaking," Paper prepared for *The Second Conference on Online Deliberation* (Stanford University, May 20-22, 2005). Paper available online at: *http://snipurl.com/fc6r*

Shane, Peter. 2005. "Turning GOLD into EPG: Lessons from the Low-Tech Democratic Experimentalism for Electronic Rulemaking and Other Ventures in Cyberdemocracy," *I/S* (1)1, 147-170.

Shellenberger, Michael and Ted Nordhaus. 2005. "The Death of Environmentalism: Global Warming Politics in a Post-Environmental World," *Grist Magazine* (January 13). Available online at: *http://www.grist.org/news/maindish/2005/01/13/doe-reprint/*

Shulman, Stuart, Jamie Callan, Eduard Hovy, and Stephen Zavestoski. 2005a. "Language Processing Technologies for Electronic Rulemaking: A Project Highlight," *Proceedings of the Fifth National Conference on Digital Government Research, May 15-18, 2005* (Digital Government Research Center). Available online at: *http://snipurl.com/f1zq*

Shulman, Stuart W., Lisa E. Thrane, and Mack C. Shelley. 2005b. "E-Rulemaking," in G. David Garson (ed.) *Handbook of Public Information Systems* (2nd ed.) (Boca Raton, FL: CRC Press), 237-254.

Shulman, Stuart, Jamie Callan, Eduard Hovy, and Stephen Zavestoski. 2004. "SGER Collaborative: A Testbed for eRulemaking Data," *Proceedings of the Fourth National Conference on Digital Government Research, May 24-26, 2004* (Digital Government Research Center). Available online at: *http://snipurl.com/f1zr*

Shulman, Stuart W. 2003. "An Experiment in Digital Government at the United States National Organic Program," *Agriculture and Human Values* 20, 3 (Fall), 253-265.

Shulman, Stuart W. 2004. "The Internet Still Might (but Probably Won't) Change Everything: Stakeholder Views on the Future of Electronic Rulemaking," A report published by the University of Pittsburgh. Available online at: *http://snipurl.com/9od6*.

Shulman, Stuart W. 2005. "eRulemaking: Issues in Current Research and Practice," *International Journal of Public Administration* 28, 7-8 (July), 621-641.

Shulman, Stuart W., David Schlosberg, Stephen Zavestoski, and David Courard-Hauri. 2003. "Electronic Rulemaking: New Frontiers in Public Participation," *Social Science Computer Review* 21(2), 162-178.

Sim, Julius and Chris C. Wright. 2005. "The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements," *Physical Therapy* 85(3), 257-268.

Stanley, J. Woody, Christopher Weare, and Juliet Musso. 2004. "Participation, Deliberative Democracy, and the Internet," in Peter Shane (ed.) *Democracy Online: The Prospects for Political Renewal Through the Internet*. (New York: Routledge), 167-179.

Suen, Hoi K. and Donald Ary. 1989. *Analyzing Quantitative Behavioral Observation Data* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Von Eye, Alexander and Eun Young Mun. 2005. *Analyzing Rater Agreement: Manifest Variable Methods* (Mahwah, New Jersey: Lawrence Erlbaum Associates).

Received: 03/07/2006
Revised: 05/10/2006
Accepted: 07/21/2006

doi:10.1300/J399v03n03_03

APPENDIX A

Original Mercury Code List (Including Minor Codes)

- *Comment on a Comment:* Text that refers to another comment submitted during the public comment process.
- *Comment on a Position:* Text that refers to a position held by an NGO, group, or citizen, BUT NOT explicitly noted as found within the docket.
- *Difficult to Code:* Text that seems not to fit anywhere, but which also seems significant, or which in some way blurs the boundaries between existing codes.
- *Disappointment:* Text that reflects the disappointment in current or future government decisions.
- *Disrespect:* Text in which the substance or tone of the comment demonstrates disrespect for another position, person, group, or comment.
- *Doomsayer:* Text that argues in the "worst case scenario" mode.
- *Economic:* Text that uses an economic rationale to make a claim.
- *Expertise:* Text that invokes an earned right to call oneself an expert (e.g., an advanced degree or job training).
- *External Authority:* Text that gives as a reason for holding an opinion that it is the view of some authority such as a trusted person, organization, religion, science, etc.
- *Good Quote:* Text that is demonstrative of the meaning of a code, the nature of the process, or which is otherwise just so interesting or funny.
- *Information in Docket (not comments):* Text reflecting that the commenter has read, and is responding to specific information in the docket BUT NOT another comment.
- *Legal:* Text that cites a legal basis to make a claim.

- *Personal Experience:* Text that invokes personal knowledge, experience, or narrative as the basis for a claim.
- *Proposal:* Text that makes a suggestion for a specific new policy or change in an existing policy.
- *Public Health and Safety:* Text shows concern for public health and/or safety.
- *Science and Technology:* Text that points to scientific or technical knowledge.
- *Social Values:* Text that invokes social values to make a claim.
- *Suspicion-Corruption:* Text that reveals a commenter is suspicious of one or more aspects/actors in the rulemaking.
- *Trust:* Text that reflects the presence of trust in government to make decisions.
- *Unique Text in a Form Letter:* Text that is suspected of being added-on to a standard form letter.

APPENDIX B

Mercury Families (Document Level)

- *Exact Duplicate:* A document that is exactly like another document, with only date, header, or signature variation.
- *Near Duplicate:* A document that is like another document, but with unique text added.
- *Suspected Duplicate:* A document that appears to be a duplicate, but the coder cannot locate a match in their bundle.
- *Off-Topic:* A document that is for another rulemaking, spam, or otherwise off the topic of the proposed mercury regulation.