

# Equality of Participation Online Versus Face to Face: An Analysis of the Community Forum Deliberative Methods Demonstration

Eric Showers and Nathan Tindall, Todd Davies\*

Center for the Study of Language and Information  
Stanford University, Stanford, California, USA  
{eshowers, ntindall, davies, lncs}@stanford.edu

**Abstract.** Public deliberation involves informed discussion by groups of citizens, representing a general public. Such groups are sometimes convened by decision makers or nongovernmental organizations as inputs to public policy. These groups have traditionally met face to face (F2F), requiring considerable time and expense. Online deliberation environments may provide a more cost-effective and/or less inhibiting environment for public participation. But do online deliberation methods (e.g. discussion boards or Internet-enhanced teleconferences) bias participation toward certain individuals or demographic groups? We compare F2F versus online contribution levels of participants in a large-scale, random assignment, U.S. deliberation experiment that allows for within-participants and cross-modal comparisons. For English speaking adults who were required to have Internet access as a condition of participation, we find no negative effects of online modes on *equality of participation* (EoP) related to gender, age, or educational level. An asynchronous discussion board/forum appears to have improved EoP for gender relative to F2F discussion. The data suggest a dampening effect of online environments on black participants, as well as amplification for white participants. Synchronous online voice communication EoP is on par with F2F across individuals (measured by Gini index). But individual-level EoP is much lower in the online forum, and greater online forum participation predicts greater F2F participation for individuals. Measured rates of participation are compared to self-reported experiences, and other findings are discussed.

**Keywords:** public deliberation · participation equality · discursive equality · e-participation

## Introduction

The efficacy of face to face deliberation has been the subject of much discussion in the academic literature. According to some authors, it leads to better decision making and

---

\* This research was supported by the Agency for Healthcare Research and Quality (AHRQ) under contract #HHSA290201000005C, through the American Institutes for Research (AIR), subcontract #641.02916. Additional funding was provided by the Vice Provost for Undergraduate Education at Stanford University. We thank AIR Staff members Coretta Mallery, Rikki Mangrum, Maureen Maurer, Manshu Yang, and Mark Rosenberg for helpful discussions and aid in obtaining the data. The views expressed in this paper are those of the authors, and have not been approved by AHRQ or AIR.

allows for a greater degree of public agency [6]. Others, however, claim that it is at best a waste of time, and at worst that it leads to bad decision making [24]. Beyond discussions of effectiveness, many authors focus on specific elements of deliberation that might prove problematic or worth looking at — the most important for the purposes of this paper having to do with the marginalization of participants according to race or gender. Some of these hypothesized effects are quantitative and can be easily identified or tested for — that men speak more than women in deliberative sessions, for example [13], [27], and [20]. Other phenomena, such as domination or the idea that some demographic groups might be less likely to deliberate or have less influence, are harder to examine quantitatively [1], [24]. Echoes of these phenomena, however, might be found in the measurable quantity of an individual's contributions. Much of this literature deals with face to face deliberation. We apply such methods to online participation as well.

As Internet access becomes more widespread and allows more users to make their voices heard, its potential as a tool for public deliberation cannot be overlooked. There is already a substantial body of literature discussing the Internet's capacity for use in this regard, e.g. [8], [12], [23], [31]. This literature is particularly concerned with the ability to eliminate some of the inequalities present in face to face deliberation [25], and cites elements of online deliberation such as anonymity and remoteness as potential benefits. There are detractors as well, however, who cite issues with online deliberation including a potential lack of respect among participants and lack of Internet access among certain groups [2]. Another dimension that must be considered is facilitation style, which itself can impact the proceedings [29]. But while the literature often discusses online and face to face (F2F) deliberation in isolation, there are few sources that provide a direct comparison between the two [22]. We aim to provide a quantitative look at examples of both kinds of deliberation in order to highlight potential differences between the deliberative modes, and to examine the effects of other variables within both modes.

In order to effectively compare online and F2F deliberation, we will examine the Agency for Healthcare Research and Quality's Community Forum Project [5],[4], which gathered together a large number of deliberative groups using different methods (as described in the next section) — one online, two offline, and one mixed. The Community Forum is beyond the scale of any controlled deliberation experiment done previously, and it sought to bring together a representative sample of the U.S. population. It is also one of the few *multiple-method* experiments that provides quantitative data on populations recruited specifically for deliberation. Our analysis is drawn from transcripts of all F2F and synchronous online meetings during the Community Forum project, archives of all online forum discussion, and records of surveys filled out by the participants that measure their knowledge, attitudes, and experience both pre- and post-deliberation.

## Community Forum Project

A five-arm randomized controlled trial was conducted between August and November 2012 by the American Institutes for Research. This Deliberative Methods Demonstration was intended to inform the Agency for Healthcare Research and Quality's research programs on public views regarding the usage of research evidence in health care decision making, and to expand the evidence base on public deliberation. The demon-

stration gathered empirical evidence about the *effectiveness* of deliberation, which has received minimal attention to date [8]. In the literature, effectiveness has been defined by the following parameters: (1) the quality of deliberative experience or discourse, (2) the changes in participants' knowledge or attitudes about the deliberative topic, (3) the changes in participants' empathy and concern for issues affecting the community at large, and (4) the impact of deliberation on decisions by the sponsoring agency.

For this Deliberative Methods Demonstration, participants were randomly assigned to one of four deliberative discussion methods, or to a reading-materials only (RMO) group. Participants were sampled from Chicago, IL, Sacramento, CA, Silver Spring, MD, and Durham, NC, where they were assigned into groups representative of the population of those areas with respect to gender, age, and ethnicity, as estimated by the U.S. Census. A total of 1,774 participants were recruited for the study, of whom 961 took part in a deliberative discussion method, and 377 were assigned to the RMO group.

The following deliberative question was posed to all participants: *Should individual patients and/or their doctors be able to make any health decisions no matter what the evidence of medical effectiveness shows, or should society ever specify some boundaries for these decisions?*

The participants were all given educational background materials to read. Those assigned to a discussion group then discussed the deliberative question in one of four distinct methods that have been advocated and used previously in prior public deliberations [5]. Additionally, some participants were only assigned reading materials. This was done to examine whether deliberation has a positive or negative impact on attitude change, and other measures of effectiveness. The main results of the study are being reported elsewhere. The methods were:

1. **Brief Citizens' Deliberation (BCD):** In this method, approximately 12 participants (the exact number varied across groups) met in person for one two hour session. The session was supported by an active facilitator. Twenty four BCD groups were convened, six in each location.
2. **Community Deliberation (CD):** This method involved two deliberative sessions among roughly 12 individuals per group. Each session was two and a half hours long, and the sessions were separated by one week. During the week between in-person sessions (CD-F2F), participants interacted through an online asynchronous discussion board (CD-Deme) [9]. Facilitation was active in the F2F sessions. There were twenty four CD groups, six at each location.
3. **Online Deliberative Polling®(ODP):** In this method, each group was convened online four times, once per week over a period of four weeks. Each meeting was a 75 minute online session, during which about 12 participants conversed through a synchronous voice interface. Facilitation was performed by students with no prior experience who were trained to intervene minimally. Twenty four ODP groups were convened.
4. **Citizens' Panel (CP):** CP involved two and a half days of deliberation. There were 24 to 30 participants in each group. Three facilitators were present when the group met. This method permitted the use of smaller breakout groups moderated by a facilitator, as well as an open space in which participants could interact without

facilitation. Facilitation in this method was otherwise active. There were four CP groups, one at each location.

5. **Reading Materials Only Control Group (RMO):** Participants assigned to the RMO intervention received educational materials via an email link. Data from these participants were not used in our investigation.

## Research Questions and Previous Findings

A few questions have been prioritized and answered in our analysis:

1. Do the medium (online versus F2F) and/or modality (e.g. speech versus text) have effects on equality of participation across demographic groups (ethnicity, gender, education, age)?
2. Do online methods differ from F2F on individual-level *equality of participation* (EoP)?
3. Do online methods differ from F2F in the effect of group size on EoP?
4. Do individuals who participate more online also participate more F2F?
5. What is the relationship between objective measures of EoP and self-reported experience?

While we were interested in broad differences between deliberative modes, of particular concern was the effect the deliberative environment had on the contributions of individuals based on their demographic. Some literature claims, for example, that women say less than men online, e.g. [10], [15], [16], [17], [28]. Does online deliberation bias contributions in favor of male participants? Other authors emphasize online divides related to race/ethnicity [19], educational level [19], and/or youth, e.g. [10], [28]. In terms of ethnicity, whites and males have been reported to say more than any other group in F2F deliberative settings as well [21], [24]. On the other hand, multiple studies of F2F deliberation have found that women speak as much as, or more than, men in these offline settings [11], [26], [30]. Are online settings different? Examining the quantitative data from the online sessions could help answer these questions.

Some literature shows that group size has an effect on F2F deliberation, and our aim was to use the vast quantity of data that the Community Forum Project collected to map that effect across its F2F and online modes. Because group size is less salient in online settings, these data provide a unique opportunity to test the hypothesis under different conditions. Finally, although the value of equality in group deliberation brings forth varying opinions in scholars, e.g. [14], [24], and [25], more unequal systems seem less desirable in cases such as public deliberation where a diversity of voices is a commonly agreed goal.

## Methods

The present study utilizes data generated in the AHRQ Community Forum Deliberative Methods Demonstration [5],[4], but this study was not conceived prior to the design of the Community Forum experiment. An optimal design for the present study would

have an online forum-only group, allowing a more pure comparison between online asynchronous text forums and the other methods. The lack of such a condition reflects limitations in the budget and aims of the Community Forum project, but we believe that much can be learned by creatively exploring the data that *were* produced.

Each deliberative session was transcribed from audio and/or video recordings. The transcripts included introductions, group level discussion, and contributions by the facilitators and experts. They typically excluded small group discussions if the method featured elements that required participants to break into smaller groups (i.e., in the CP method).

These transcripts were scraped using a C++ script in order to associate each participant with their contributions. A contribution was defined to be a paragraph of text that had an identifiable speaker and an associated body. The body was at least one word long, and had no upper bound. 75,115 total contributions were scraped from the transcript files. Of these contributions, 70,492 were able to be linked by participant ID to a session participant in the Community Forum study or to the facilitator. The contributions that were discarded had speakers that were unable to be associated with a participant in that deliberative session, and most commonly had anonymous IDs that indicated that the speaker was unable to be ascertained from the recordings by the transcription firm (i.e. Man 1, Respondent, Unknown, etc.). These 70,492 contributions were used in data analysis (93.8% of total scraped). Despite an indeterminate level of omission noise in the transcripts, the size of the data set provides us with bountiful information about how individuals contribute in group discussion.

For each contribution, the number of words it comprised was tabulated. From these data, the *frequency*, *volume*, and *average contribution length (ACL)* were calculated for each individual in the deliberative sessions. The frequency of contribution was calculated by dividing an individual's number of spoken continuous contributions by the total number of contributions spoken in the session. The volume of contributed words for an individual was calculated by dividing the total number of words that an individual spoke by the total number of words that were spoken in the session by all participants. The average contribution length was calculated by taking the total number of words that an individual spoke and dividing it by their number of contributions. Measuring frequency and volume as percentages was necessary to perform analysis across methods due to variation in deliberation duration and group size.

The following were considered independent variables, as self reported by each participant: age, gender, education, and race/ethnicity (Hispanic, Native American, Asian or Pacific Islander, Black or African American, White, Other). Education was self-reported as one of eight categories, increasing from "less than high school graduate" to "more than 4-year college graduate." Individuals could indicate more than one race/ethnicity.

Table 1 shows the number of transcript files that were scraped from each method, the range in attendance for sessions, and the demographic makeup of the people who participated. For the analysis, each session participant was treated equally, and thus for the methods that featured multiple sessions, the same individual is being counted more than once. We made this decision because we were primarily interested in the contribution patterns of individuals once they were participating in a controlled deliberative

**Table 1. Mean Values: Demographic Data**

Method	Groups	Sessions	Individuals	Size Range	Avg. Size	Avg. Age	Fem-Prop	Education	Hispanic	Native	Asian	Black	White	Other
BCD	24	1	309	9-14	13.0	46.9	0.55	5.49	0.13	0.02	0.03	0.27	0.60	0.10
CD	48	2	292	7-13	11.8	47.5	0.55	5.65	0.11	0.01	0.02	0.33	0.56	0.09
CP	12	3	98	20-28	24.3	48.5	0.57	5.39	0.10	0.00	0.01	0.47	0.43	0.10
ODP	72	4	262	5-12	9.5	45.6	0.52	5.87	0.11	0.01	0.01	0.25	0.64	0.11

setting, and were less interested in the changes in contribution that occur over time. Time sensitive changes in contribution deserve further study; however, *our analysis focuses on the behavior of individuals once they enter a deliberative setting.*

Analysis was performed across methods, across media (online/offline), and by looking at isolated subpopulations in order to investigate the behavior of different ethnic and gender subgroups. Deliberative experience surveys were also administered. After completing their deliberative method, participants were asked to answer the following questions (among many others) on a scale of disagree strongly to agree strongly: DE 05: "Some people in the group spoke a lot more than others;" DE 06: "Some people in the group spoke too much;" DE 07: "Some people in the group barely spoke at all;" DE 08: "I spoke as much as I wanted to in the group." The equality factor reported was calculated by AIR to have a Cronbach's alpha value of 0.64, as a function of questions DE 05, DE 06, and DE 07, found by exploratory factor analysis [5].

## Findings

We divide the findings into three parts. The first compares the ODP (synchronous voice) data with the three other F2F methods. The second compares participants in the CD group who posted on the online forum (asynchronous text) with those who did not. And the third reports findings that speak to EoP across deliberative modes.

### Synchronous Voice vs. Face to Face

Tables 2 through 4 show the frequency, volume, and average contribution length correlations with different independent variables across all five environments: the ODP and CD-Forum (the online environments) and the F2F component of CD (which we will call CD-F2F), BCD, and CP environments. Significant negative effects for attendance (group size) were found with respect to frequency and volume across all four methods but no effects with respect to average contribution length were found to be significant. Correlations and p-values were as follows.

- BCD: frequency,  $\rho = -0.151$ ,  $p < 0.015$ ; volume,  $\rho = -0.140$ ,  $p < 0.02$ .
- CD-F2F: frequency,  $\rho = -0.222$ ,  $p < 0.001$ ; volume,  $\rho = -0.196$ ,  $p < 0.001$ .
- CP: frequency,  $\rho = -0.134$ ,  $p < 0.03$ ; volume,  $\rho = -0.121$ ,  $p < 0.05$ .
- ODP: frequency,  $\rho = -0.304$ ,  $p < 0.001$ ; volume,  $\rho = -0.252$ ,  $p < 0.001$ .

Significant positive effects for age were found across the various methods as well.

- BCD: frequency,  $\rho = 0.211$ ,  $p < 0.001$ .

**Table 2. Frequency Correlations**

Mode	Method	Size	Age	Gender	Education	Hispanic	Native	Asian	Black	White	Other
Online	ODP	-0.304***	0.294**	0.015	0.042	-0.073	0.009	-0.050	-0.020	0.106**	-0.129***
	CD-Forum	0.010	0.106	-0.016	0.115	0.030	-0.081	-0.080	-0.033	0.097	-0.061
F2F	CD	-0.222***	0.107*	-0.041	0.101*	-0.064	0.002	-0.018	0.017	0.010	-0.059
	BCD	-0.150*	0.211***	-0.081	0.119	0.038	0.062	-0.068	-0.186**	0.190***	0.000
	CP	-0.134*	0.134*	0.012	0.118	-0.128*	—	-0.062	-0.127*	0.166**	-0.070

**Table 3. Volume Correlations**

Mode	Method	Size	Age	Gender	Education	Hispanic	Native	Asian	Black	White	Other
Online	ODP	-0.252***	0.203***	-0.0156	0.160***	-0.066	0.037	-0.018	-0.134***	0.191***	-0.127***
	CD-Forum	0.008	0.112	-0.029	0.145	0.006	-0.072	-0.074	-0.055	0.122*	-0.068
F2F	CD	-0.196***	0.047	-0.055	0.138*	-0.045	0.022	-0.032	0.030	0.0035	-0.035
	BCD	-0.140*	0.087*	-0.147*	0.136	0.041	0.038	-0.065	-0.127*	0.147*	-0.004
	CP	-0.121*	0.023	0.001	0.185**	-0.011	—	-0.065	0.031	0.006	0.014

**Table 4. Average Contribution Length Correlations**

Mode	Method	Size	Age	Gender	Education	Hispanic	Native	Asian	Black	White	Other
Online	ODP	0.060	0.063	-0.056	0.167***	0.023	0.023	0.087*	-0.149***	0.116**	-0.032
	CD-Forum	0.048	0.091	-0.056	0.151*	0.002	-0.085	-0.075	-0.030	0.094	-0.055
F2F	CD	-0.054	-0.166***	0.032	0.092*	0.043	0.175***	-0.044	0.032	-0.063	0.121**
	BCD	0.032	-0.161**	-0.117	0.156**	0.032	-0.099	0.019	0.084	-0.047	-0.033
	CP	0.075	-0.112	-0.083	0.129**	0.179	—	-0.043	0.024	-0.132*	0.163**

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$

- CD-F2F: frequency,  $\rho = 0.107, p < 0.001$ ; average length,  $\rho = -0.166, p < 0.001$ .
- CP: frequency,  $\rho = 0.134, p < 0.02$ , volume,  $\rho = 0.023, p < 0.05$ .
- ODP: frequency,  $\rho = 0.249, p < 0.001$ ; volume,  $\rho = 0.204, p < 0.001$ .

With respect to education, a positive relationship between contribution and self-reported education was found.

- BCD: volume,  $\rho = 0.136, p < 0.03$ ; average length,  $\rho = 0.156, p < 0.01$
- CD-F2F: frequency,  $\rho = 0.101, p < 0.03$ ; volume,  $\rho = 0.139, p < 0.002$ ; average length,  $\rho = 0.092, p < 0.04$ .
- CP: volume,  $\rho = 0.185, p < 0.02$ ; average length,  $\rho = 0.129, p < 0.033$ .
- ODP: volume,  $\rho = 0.160, p < 0.001$ ; average length,  $\rho = 0.167, p < 0.001$ .

For the ODP, CD-F2F, and CP conditions, no significant effect was found between gender and contribution frequency, volume, or average length. However, in the BCD condition, female identification had a significant negative correlation with volume ( $\rho = -0.147, p < 0.02$ ).

In the ODP condition, a participant indicating that they were white had a positive, significant correlation with all contribution metrics (frequency,  $\rho = 0.107, p < 0.001$ ; volume,  $\rho = 0.191, p < 0.001$ ; average length,  $\rho = 0.116, p < 0.01$ ), while black identification had a negative correlation with volume ( $\rho = -0.134, p < 0.001$ ). A similar trend was found in the BCD condition, where white identification had a positive, significant

**Table 5. Mean Values: Posters vs. Nonposters**

Subset	Avg. Age	Fem-Prop	Education	Hispanic	Native	Asian	Black	White	Other	Frequency	Volume	Avg. Length
Posters	50.0	0.58	5.74	0.12	0	0.01	0.35	0.56	0.08	0.10	0.10	35.8
Nonposters	46.7	0.52	5.52	0.10	0.03	0.03	0.30	0.55	0.118	0.084	0.085	41.1
<i>p</i> -value	0.31	0.15	0.25	0.51	0.01*	0.07	0.24	0.83	0.12	0.01**	0.03*	0.01**

correlation with frequency and volume (frequency,  $\rho = 0.190$ ,  $p < 0.001$ , volume:  $\rho = 0.147$ ,  $p < 0.02$ ), while answering the ethnicity question with "Black or African American" had negative contribution correlations (frequency,  $\rho = -0.186$ ,  $p < 0.002$ ; volume,  $\rho = -0.127$ ,  $p < 0.05$ ). For the CP condition, answering ethnicity with "White" had a positive, significant correlation with frequency ( $\rho = 0.166$ ,  $p < 0.01$ ), while black identification had a negative, significant correlation with frequency ( $\rho = -0.127$ ,  $p < 0.05$ ). No significant effects with respect to ethnicity were found for the CD-F2F method. Identifying as "Hispanic", "Native American", or "Asian or Pacific Islander" showed no systematic correlations with contribution measures across methods.

### Findings Within the Citizens' Deliberation Hybrid Method

**Posters vs. nonposters.** Table 5 compares those who posted in the online forum of the CD method. The frequency, volume, and average length figures given there are for each group's average-member contributions in the F2F sessions of CD. Tables 6 through 8 compare the poster and nonposter groups in CD both online and F2F in terms of the demographic variables and sizes of the groups in which they were participating,

Posters and nonposters showed similar, significant effects with respect to attendance:

- Posters: frequency,  $\rho = -0.234$ ,  $p < 0.001$ ; volume,  $\rho = -0.170$ ,  $p < 0.002$ .
- Nonposters: frequency,  $\rho = -0.225$ ,  $p < 0.001$ ; volume,  $\rho = -0.248$ ,  $p < 0.001$ .

Posters showed positive effects for frequency and volume with respect to age, but nonposters showed no effect, a discrepancy from the findings of the other groups. Nonposters and posters shared significant negative correlations for average contribution length with respect to age, however.

- Posters: frequency,  $\rho = 0.177$ ,  $p < 0.001$ ; volume,  $\rho = 0.119$ ,  $p < 0.002$ ; average length,  $\rho = -0.143$ ,  $p < 0.011$ .
- Nonposters: average length,  $\rho = -0.185$ ,  $p < 0.008$ .

Posters showed positive, significant correlations with respect to education only for average contribution length, while nonposters showed positive, significant correlations with respect to education for frequency and volume:

- Posters: average length,  $\rho = 0.126$ ,  $p < 0.01$ .
- Nonposters: frequency,  $\rho = 0.165$ ,  $p < 0.02$ ; volume,  $\rho = 0.197$ ,  $p < 0.005$ .

No systematic, significant correlations were found for ethnicity or gender among the poster and nonposter groups, with the exception that female identification had a



**Table 6. Frequency Correlations**

Mode	Subset	Size	Age	Gender	Education	Hispanic	Native Asian	Black	White	Other
Online	Posters	-0.165*	0.1134	-0.069	0.136	0.006	—	-0.063	-0.104	0.158* -0.034
F2F	Posters	-0.234***	0.177**	-0.100	0.050	-0.011	—	-0.122*	-0.047	0.059 0.018
	Nonposters	-0.225***	0.003	0.016	0.165*	-0.166*	0.030	0.074	0.122	-0.082 -0.138

**Table 7. Volume Correlations**

Mode	Subset	Size	Age	Gender	Education	Hispanic	Native Asian	Black	White	Other
Online	Posters	-0.138	0.124	-0.080	0.181*	-0.026	—	-0.061	-0.126	0.187* -0.051
F2F	Posters	-0.170**	0.119*	-0.135*	0.094	0.000	—	-0.115	-0.030	0.046 0.034
	Nonposters	-0.249***	-0.054	0.038	0.197**	-0.131	0.055	0.038	0.119	-0.070 -0.106

**Table 8. Average Contribution Length Correlations**

Mode	Subset	Size	Age	Gender	Education	Hispanic	Native Asian	Black	White	Other
Online	Posters	-0.082	0.0885	-0.140	0.201**	-0.041	—	-0.044	-0.107	0.159* -0.02
F2F	Posters	0.034	-0.142*	-0.005	0.126*	0.102	—	-0.091	0.017	-0.25 0.124*
	Nonposters	-0.149*	-0.185**	0.103	0.065	-0.024	0.229	-0.039	0.061	-0.104 0.104

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$

significant, negative correlation with contributed volume in the F2F session ( $\rho = 0.197$ ,  $p < 0.05$ ).

In comparing the total F2F contributions of posters and nonposters (Table 5), posters' contributions were of significantly higher frequency ( $p < 0.005$ ), significantly higher volume ( $p < 0.03$ ), and their average contribution lengths were significantly less ( $p < 0.01$ ).

**Face to face vs. asynchronous text (forum).** In the F2F component of CD, group size had a statistically significant negative correlation with both frequency ( $\rho = -0.222$ ,  $p < 0.001$ ) and volume ( $\rho = -0.196$ ,  $p < 0.001$ ), but no significant impact on average length. These results were mirrored in the online component (frequency,  $\rho = -0.165$ ,  $p < 0.05$ ; volume,  $\rho = -0.138$ ,  $p < 0.08$ ) although the effect was weakened. (See Tables 6, 7, and 8 for this subsection.)

There were no significant age effects in the online case, but in the F2F sessions, frequency was positively correlated with age ( $\rho = 0.107$ ,  $p < 0.02$ ), while average length was negatively correlated  $\rho = -0.166$ ,  $p < 0.001$ ). There were no significant effects for gender in either medium. Educational level was positively correlated with all contribution measures both online (frequency,  $\rho = 0.136$ ,  $p < 0.08$ ; volume,  $\rho = 0.181$ ,  $p < 0.02$ ; average length,  $\rho = 0.201$ ,  $p < 0.01$ ) and in the F2F sessions (frequency,  $\rho = 0.100$ ,  $p < 0.03$ ; volume,  $\rho = 0.139$ ,  $p < 0.01$ ; average length,  $\rho = 0.092$ ,  $p < 0.05$ ), with slightly stronger effects online.

There were no significant race/ethnicity effects among the F2F participants, but white identification had a positive, significant correlation with all metrics in the online

case (frequency,  $\rho = 0.158$ ,  $p < 0.05$ ; volume,  $\rho = 0.187$ ,  $p < 0.02$ ; average length,  $\rho = 0.159$ ,  $p < 0.05$ ).

**Face to face (posters only) vs. forum in CD.** As shown in Tables 6 through 8, we also examined differences between the behavior of those who posted online and spoke offline in CD, in order to examine if the change in medium would impact individuals' contribution rates. Group size effects were consistent with the other methods, though the effect observed in the F2F mode (frequency,  $\rho = -0.234$ ,  $p < 0.001$ ; volume,  $\rho = -0.169$ ,  $p < 0.002$ ) was much stronger than in the asynchronous forum setting (frequency,  $\rho = -0.153$ ,  $p < 0.04$ ; volume,  $\rho = -0.125$ ,  $p < 0.08$ ).

Although no significant age effects were found in the online medium, the effect was significant across all metrics in the F2F setting, (frequency,  $\rho = 0.176$ ,  $p < 0.002$ ; volume,  $\rho = 0.119$ ,  $p < 0.04$ ; average length,  $\rho = -0.143$ ,  $p < 0.02$ ).

In the F2F condition, education had a positive and significant effect on average contribution length ( $\rho = 0.126$ ,  $p < 0.02$ ), and was similar online (frequency,  $\rho = 0.136$ ,  $p < 0.08$ ; volume,  $\rho = 0.181$ ,  $p < 0.02$ ; average length,  $\rho = 0.201$ ,  $p < 0.01$ ).

Among posters, women contributed less in the F2F setting (frequency,  $\rho = -0.100$ ,  $p < 0.08$ ; volume,  $\rho = -0.135$ ,  $p < 0.04$ ), though no significant gender effects were observed in the online setting.

With respect to ethnicity, no systematic effects were observed in the F2F case. However, significant effects were observed for white posters on the online forum, who posted more than those who were nonwhite (frequency,  $\rho = 0.158$ ,  $p < 0.05$ ; volume,  $\rho = 0.187$ ,  $p < 0.02$ ; average length,  $\rho = 0.159$ ,  $p < 0.05$ ).

### Equality of Participation Across Individuals

The Gini index is calculated based on the cumulative distribution function for each participation measure. Although the most common application of the Gini index is its use as a measure of income inequality in a given nation, it also can be used as a measure of inequality in a data set. In this context the Gini index ranges from 0, representing complete equality, to 1, representing complete inequality. Gini indices were calculated for each session, and the values analyzed for each medium, in order to investigate EoP differences across methods.

The Gini index was calculated by the following formula, which fulfills the Transfer Principle of Inequality [18], where  $X_i$  is the amount that the person contributed and  $P_i$  is the contribution rank of person  $i$  such that the person who contributed most receives a rank of 1 and the person who contributed least a rank of  $N$ :

$$G = \frac{N+1}{N-1} - \frac{2}{N(N-1)\bar{x}} \sum_{i=1}^n P_i X_i$$

In comparing the synchronous voice method (ODP) against the other methods (Table 9), statistically significant differences were found for frequency between ODP and BCD ( $p < 0.05$ ), and between ODP and the (Deme) Forum ( $p < 0.05$ ). Significant differences for volume were found between ODP and both BCD and CD-F2F ( $p <$

**Table 9. Gini indices across methods**

Method	Frequency	Volume	Average Length
BCD	0.329	0.351	0.204
CD-F2F	0.335	0.368	0.214
CP	0.400	0.448	0.283
ODP	0.362	0.439	0.279
CD-Forum	0.754	0.702	0.556

**Table 10. Gini indices among posters across mediums**

Method	Frequency	Volume	Average Length
Forum	0.316	0.467	0.302
Face to face	0.322	0.345	0.203

0.001). Additionally, significant differences for average contribution length were found between ODP and both BCD and CD-F2F ( $p < 0.001$ ). For volume, ODP ( $G = 0.439$ ) and CP ( $G = 0.448$ ) reported the highest Gini indices, with the other F2F methods showing more modest, yet still fairly high, coefficients (BCD:  $G = 0.351$ ; CD-F2F:  $G = 0.368$ ). The Gini indices for the online forum when including all participants in the CD method were dramatically greater than for all the other methods, indicating, perhaps not surprisingly, that an optional online forum draws a more limited set of participants.

Gini indices for each metric were also calculated based on the F2F contributions of the people who posted (Table 10). Among those who posted, a statistically significant difference between online ( $G = 0.467$ ) and F2F ( $G = 0.345$ ) media is prominent for volume ( $p < 0.001$ ). Even among those who choose to participate in an online forum, there appears to be less EoP for volume and average length (thought not frequency).

**Self reported experiences regarding equality of participation.** Tables 11-13 relate the results of the post-deliberative experience survey to our measures of EoP. Table 11 shows that participants rated CD (F2F) and BCD the most equal of the methods. High scores on the equality factor in the last column indicate more perceived equality as measured by the combination of questions DE 05, DE 06, and DE 07 that the equality factor represents (see previous section on Methods), with CP being the least perceived equal, and ODP falling in between. The one online deliberative method (ODP) that participants rated for equality was thus neither the most nor the least unequal when it comes to participation in the eyes of participants. Question DE 08 asked participants to express agreement or disagreement with the statement, "I spoke as much as I wanted to in the group." Participants rated CD-F2F significantly higher on DE 08 than they rated BCD or ODP, but once again ODP was at least on a par with one of the F2F methods (BCD) in promoting participants' satisfaction with the amount that they individually spoke. An interesting comparison is with the measured Gini indices for each method (Table 9). The subjective equality factors (Table 11) roughly mirror the pattern of Gini indices across the four rated methods, with BCD and CD(-F2F) scoring as the most

**Table 11. Self Reported Equality: Across Methods**

Method	DE 05	DE 06	DE 07	DE 08	Equality Factor
BCD	2.818**	2.122	2.663**	3.237	2.465***
CD-F2F	2.877*	2.076**	2.637***	3.447***	2.471***
CP	3.106*	2.661***	2.854	3.344*	2.143***
ODP	2.972	2.211	2.799	3.259	2.339

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$  w.r.t. ODP value for this question or factor

**Table 12. Self Reported Equality: Posters vs. Nonposters**

Method	DE 05	DE 06	DE 07	DE 08	Equality Factor
Posters	2.929	2.071	2.698	3.492	2.434
Nonposters	2.790	2.084	2.532	3.379	2.531
<i>p</i> -value	0.020*	0.830	0.011*	0.029*	0.039*

\* $p < .05$

**Table 13. Experience Surveys, by race**

Method	Race	DE 05	DE 06	DE 07	DE 08	Equality Factor
BCD	White	2.795	2.145	2.735***	3.205	2.436
	Black	2.841	2.057	2.500	3.329	2.533
CD-F2F	White	2.943**	2.114	2.790***	3.448	2.384***
	Black	2.771	2.036	2.355	2.476	2.612
CP	White	3.128	2.932***	3.051***	3.377	1.963***
	Black	3.150	2.370	2.772	3.307	2.236
ODP	White	3.048*	2.279***	2.881***	3.231**	2.264***
	Black	2.887	2.013	2.596	3.371	2.501
CD-Posters	White	2.988**	2.091	2.841***	3.438*	2.360***
	Black	2.776	2.056	2.402	3.598	2.588
CD-Nonposters	White	2.867	2.152	2.705***	3.467*	2.425**
	Black	2.754	2.000	2.246	3.263	2.667

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$  between races

equal on all three Gini measures and also on the subjective equality factor, CP scoring the least equal on all, and ODP scoring in the middle on all.

Table 12 shows, interestingly, that posters in the CD-Deme forum rated overall equality significantly lower in the CD-F2F environment than did nonposters, though posters were more satisfied that they personally said what they wanted to (DE 08). We saw in Table 5 that posters exceeded the contributions of nonposters in the F2F sessions of CD by all three objective metrics, and their reported satisfactions can be reconciled with this fact. Posters appeared aware that they got more than their share of speaking in during the CD-F2F sessions, perhaps leading them to feel satisfied with their own participation but less satisfied that the process produced equal participation.

Table 13 compares self-identified white and black participants on the experience survey questions related to equality of participation. It shows that, in all but one of

**Table 14. Volume normalized by multiple regression on age and education**

Metric	Race	BCD	CD-F2F	CP	ODP	CD-Deme	CD-F2F*
Avg. age	White	48.599	51.021	52.949	47.845	52.226	51.708
	Black	44.689	43.174	45.789	41.448	43.931	43.387
Avg. education	White	5.749	5.959	5.692	6.143	6.108	6.045
	Black	4.689	5.122	4.960	5.157	5.224	5.270
Avg. volume (%)	White	9.272	9.250	4.405	12.567	16.793	9.985
	Black	7.512	9.496	4.258	9.181	12.082	9.489
Pred. volume (%)	White	9.189	8.854	4.168	12.216	16.418	9.955
	Black	8.031	9.533	4.238	10.256	13.026	9.512

\* Posters only

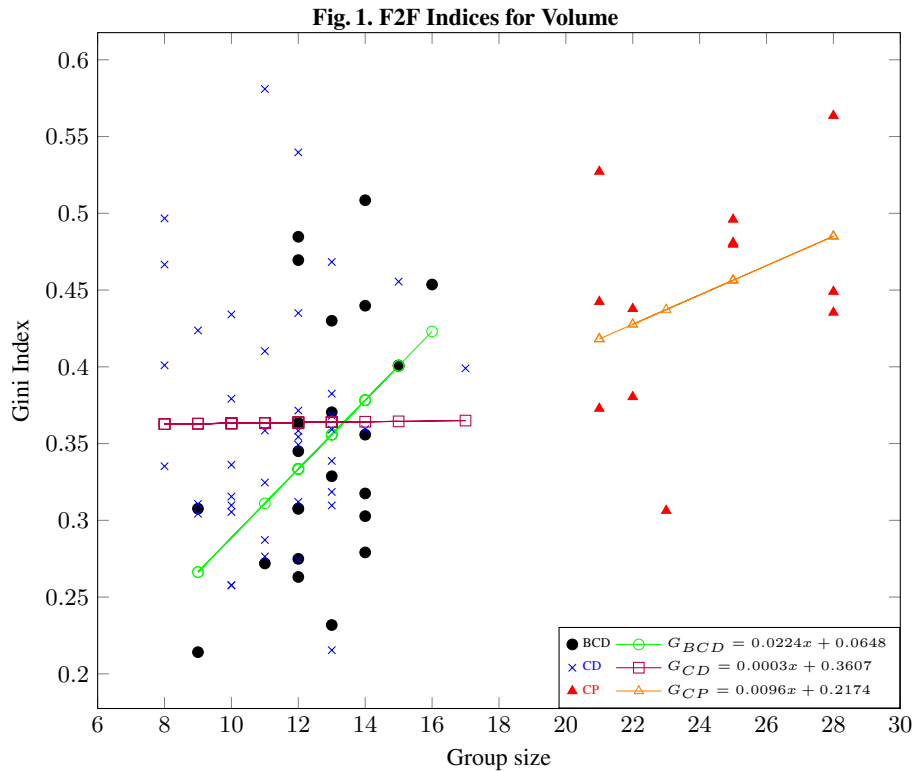
the methods (BCD), white participants rated the equality of the method significantly lower than did black participants, despite the fact that black identification predicted lower volume of participation in all but one of the method groups (see Table 14, “Avg. volume” and “Pred. volume” – the latter is normalized for age and educational level).

#### Individual-level equality of participation and group size online versus face to face.

The Gini index is a measure of how concentrated participation is across individuals (the higher the Gini, the more participation is dominated by a subgroup of participants). A natural question to ask is what effect the size of the group has on this measure. As can be seen in Fig. 1, the Gini index goes up substantially as group size increases in both the BCD and CP methods, but is unaffected by group size in the CD method. Fig. 2 shows the relationship between Gini indices and group size for the online environments: it is either flat or slightly negative for both ODP (synchronous voice) and CD-Deme (the asynchronous text forum). This provides some evidence that the two online methods each scale well, at least within the observed ranges (7-12 and 8-17 participants, respectively, for the ODP and CD groups). Adding more participants within these ranges does not seem to make participation more unequal across individuals in the two online methods, nor in the CD-F2F method. But adding participants does seem to reduce EoP in two of the F2F methods: BCD and CP.

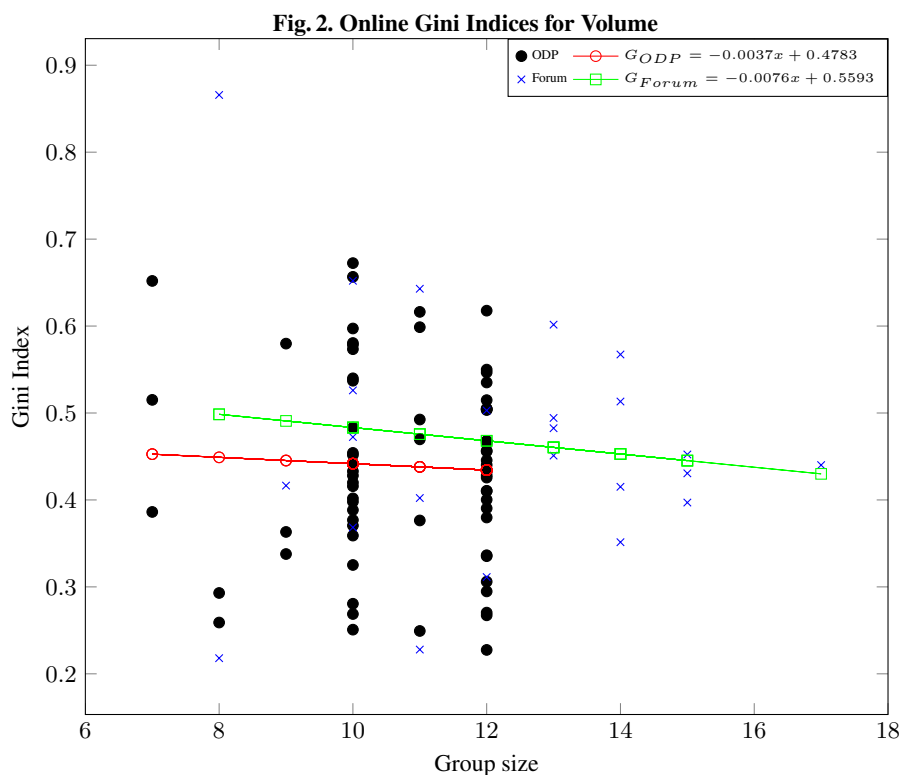
## Discussion

Some literature argues that women are less likely to participate online than are men, e.g. [3], [2], though women may be more likely to participate equally with men online than offline [22]. We found no significant negative effects on EoP for women across methods, with the exception of the BCD method, which favors men in volume. This conclusion deviates from the sizeable body of literature arguing that women speak less F2F [20], [7], [22], and from the claim that women are less active in online contributing [2], [22], in agreement with the idea that online environments do not adversely impact gender EoP. One explanation for the F2F equality of contribution is that all of the F2F methods were facilitated, and there is evidence to support that facilitation eliminates the worst of the gender gap in deliberation [29]. The discrepancy in our findings lies with the BCD method, in which female identification negatively correlated with all contribution



measures. The only difference between BCD and the other F2F methods was that BCD used a male facilitator for half of the groups (rather than a female for all), but women were not significantly more inhibited under the male facilitator than under the female one. The data do not suggest a compelling reason for this correlation, but it is possible that it has to do with different facilitation methods. Despite this discrepancy, this implies that the difference of medium (online vs. offline) is not causing the difference in and of itself. Indeed, in the CD condition we saw that female online forum posters participated equally with men, but the same women did not do so F2F. These findings agree with some other studies involving online deliberation, most notably another study in which participants deliberated on healthcare issues [23], but they are unique in being drawn from a within-group study. Previous examinations of online deliberation even when compared directly with F2F deliberation, have not used the same group that participated F2F when tracking online contributions.

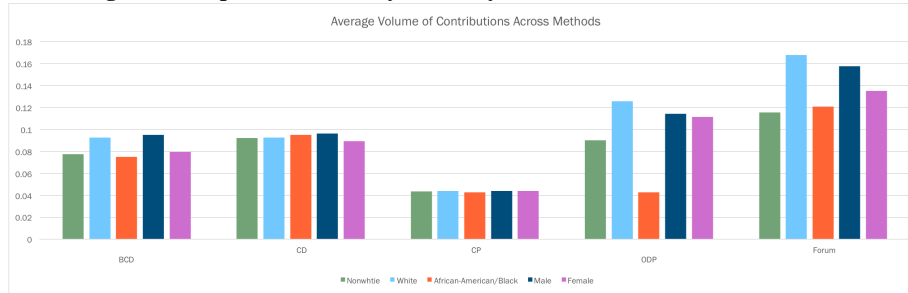
In most of the methods, and most visibly online, there was a significant positive correlation between white identification and volume/frequency, and a significant negative correlation between black identification and the same measures. This was most prevalent in the online methods — ODP and the forum — where facilitation was the least present. An effective illustration of the difference between the online and F2F environments is the graph in Figure 3, which shows relatively even participation levels across



ethnicities in the three F2F environments, but noticeable differences in both ODP and the CD-Deme (Forum) setting. ODP is unique in that there did not seem to be a tradeoff for white participants between volume/frequency and ACL, and the correlation with all three measures was positive. These results are consistent with other findings for gender, in which the gender gap is eliminated via facilitation [29].

Noticing that age and education had consistent positive relationships with contribution measures for all metrics, a multiple regression model was generated in order to investigate to what extent these factors could compensate for other discrepancies, especially between ethnicities. As shown in Table 14, the racial difference in participation is reduced when we control for age and education (compare Avg. volume, which is observed, with Pred. volume, the predicted percentages based on the white participants' averages for age and educational level). A gap persists for the two online environments: ODP and CD-Deme (Deme is the Forum component of CD), though not for the F2F environments, including CD-F2F\*, which represents just Deme forum posters in the F2F component of CD (the same populations as are represented in the CD-Deme column). This merits further study to determine whether the media difference is robust.

In our comparison within the CD method of posters and nonposters, people who posted in the forums tended to say significantly more F2F (by both frequency and volume measures) than people who did not post (Table 5). Factors such as civic participa-

**Fig. 3. Participation Volume by Ethnicity and Gender in Different Methods**

tion are effective predictors of the likelihood of citizens participating online [23], and it is likely that volume and frequency of F2F participation in the offline CD setting could also serve as significant predictors for posting online. Posters tended to have a shorter average contribution length than nonposters, but this is in line with a larger trend in the results — of a tradeoff between average contribution length and volume/frequency.

Group size, as expected, had a negative correlation with volume and frequency in all the F2F and the ODP modes of deliberation, in which the procedures were timed and there might have been pressure on individuals to allow time for others to speak. Unexpectedly, however, this correlation carried over into the forum as well. Given that there was no time pressure in the Forum, it is unclear why the volume and frequency of forum posters also decreased as group size increased.

The individual-level inequality measurements suggest a few things — first, that larger group size correlates positively with higher inequality, which explains the discrepancy between the CP method and BCD/CD methods. There is a potential interaction with intensity as well, as CD was more unequal than BCD. The online methods each were more unequal than both BCD and CD, which might be attributable to facilitation — the facilitation style of both BCD and CD were more active than that used in ODP, while the forums had experts who weighed in on conversations but were also not actively facilitating. A useful comparison is with income: a  $G$  score of 0.300 maps to some of the lowest income inequalities among countries of the world.

## Conclusions

While there have been a variety of studies of online and offline deliberation, none have as large a pool of information to work with as the Community Forum project, and as such it provides a unique opportunity to provide quantitative analysis of the difference between the two modes on a scale that has not been seen before. While this paper does not represent an exhaustive report of all the conclusions that can be gleaned from the data, some conclusions appear well-supported based on our analysis so far:

1. *Online effects on demographic groups' participation equality.* Overall, we see no consistent effects of online versus F2F participation for gender equality of participation in these data. There is evidence that some deliberation methods (e.g. the F2F BCD



method) may adversely impact female participation, independently of the offline-online dimension, and that an asynchronous forum produces higher EoP across genders than F2F discussion. For ethnicity, the online versus F2F picture is less clear, but the online settings in this study do seem to have depressed black participation relative to whites' (see Figure 3). Online deliberation appears to reduce black and increase white participation somewhat, relative to F2F, even when controlling for age and educational level (see Table 14). This provisional finding requires further investigation. Older participants appear to contribute more in volume online (see Table 3), possibly because the negative effect of age on average contribution length that we see in F2F environments does not occur online. Online environments do not appear to amplify participation inequality related to educational level, which might be a bit of a surprise (see Table 3).

2. *Online effects on individual-level participation equality.* As measured by Gini indices, synchronous voice deliberation (ODP) is on par with F2F methods for individual-level EoP (Table 9). But the optional online forum used in CD (Deme) produced much greater concentration of participation volume than did F2F methods, including the CD-F2F environment that included the same participants (Table 10).

3. *Online environments and group size effects.* Although the methods tested here are too limited to say so definitively, in this study the online environments (ODP and CD-Forum) eliminated the amplification of inequality that we saw from group size in the BCD and CP (but not CD-F2F) methods (Figures 1 and 2).

4. *Online posting as a predictor of F2F participation.* In Table 5 we saw that Forum posters in the CD method out-participated nonposters on all three contribution metrics, indicating that the tendency for an individual to participate is correlated across online and F2F contexts.

5. *Relationship of self-reported experience to measures of participation equality.* The Gini coefficients for frequency, volume, and ACL, as measures of individual-level EoP, proved to be good predictors both of each other and of the subjective equality factor (Tables 9 and 11). Interestingly, however, at the demographic level there was a more puzzling relationship. Black participants rated all but one of the methods more equal than did white participants, even when they participated less by volume than white identified participants did. The ODP method was the only pure test of subjective ratings for an online method, and, consistent with its Gini coefficients, participants rated it neither the most nor the least equal in comparison to the other (F2F) methods.

For further research, the findings related to gender could be taken in a more focused direction. Though ODP was an exercise in synchronous voice deliberation, the purpose of the forum was question-answering rather than deliberation proper. Using a method similar to CD in which the asynchronous text component were used to deliberate, rather than to share personal anecdotes and ask questions about the topic, would provide a better test of gender equality between online and offline methods. Future research might place more emphasis on individual group composition and its effects on individual contributions, to isolate the cause of demographic trends. Additionally, though outside the scope of this paper, looking at facilitator trends might prove useful — the different effects between the two BCD facilitators were significant, and merit a deeper investigation.

## Bibliography

- [1] Abdel-Monem, T., Bingham, S., Marincic, J., Tomkins, A.: Deliberation and diversity: Perceptions of small group discussions by race and ethnicity. *Small Group Research* 41(6), 746–776 (2010)
- [2] Albrecht, S.: Whose voice is heard in online deliberation?: A study of participation and representation in political debates on the internet. *Information, Community and Society* 9(1), 62–82 (2006)
- [3] Baek, Y.M., Wojcieszak, M., Carpini, M.X.D.: Online versus face-to-face deliberation: Who? why? what? with what effects? *New media & society* 14(3), 363–383 (2012)
- [4] Carman, K.L., Mallery, C., Maurer, M., Wang, G., Garfinkel, S., Yang, M., Gilmore, D., Windham, A., Ginsburg, M., Sofaer, S., et al.: Effectiveness of public deliberation methods for gathering input on issues in healthcare: Results from a randomized trial. *Social Science & Medicine* 133, 11–20 (2015)
- [5] Carman, K.L., Maurer, M., Mallery, C., Wang, G., Garfinkel, S., Richmond, J., Gilmore, D., Windham, A., Yang, M., Mangrum, R., et al.: Community forum deliberative methods demonstration: Evaluating effectiveness and eliciting public views on use of evidence (2014)
- [6] Christiano, T.: The significance of public deliberation. *Deliberative democracy: Essays on reason and politics* 243, 249 (1997)
- [7] Dahlberg, L.: Computer-mediated communication and the public sphere: A critical analysis. *Journal of Computer-Mediated Communication* 7(1), 0–0 (2001)
- [8] Davies, T., Chandler, R.: Online deliberation design: Choices, criteria, and evidence. In: *Democracy in Motion: Evaluating the Practice and Impact of Deliberative Civic Engagement*. Oxford University Press (2012)
- [9] Davies, T., Mintz, M.D., Tobin, J., Ben-Avi, N.: Document-centered discussion and decision making in the deme platform. In: *Collaboration Technologies and Systems (CTS), 2012 International Conference on*. pp. lxxv–lxxviii. IEEE (2012)
- [10] Davis, R.: *Politics online: Blogs, chatrooms, and discussion groups in American democracy*. Routledge (2005)
- [11] Dutwin, D.: The character of deliberation: Equality, argument, and the formation of public opinion. *International Journal of Public Opinion Research* 15(3), 239–264 (2003)
- [12] Fishkin, J.S.: Virtual public consultation: Prospects for internet deliberative democracy. *Online deliberation: Design, research, and practice* pp. 23–35 (2009)
- [13] Fraser, N.: Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text* pp. 56–80 (1990)
- [14] Gerber, M.: Equal partners in dialogue? participation equality in a transnational deliberative poll (europolis). *Political Studies* (2015)
- [15] Hargittai, E., Walejko, G.: The participation divide: Content creation and sharing in the digital age 1. *Information, Community and Society* 11(2), 239–256 (2008)

- [16] Harp, D., Tremayne, M.: The gendered blogosphere: Examining inequality using network and feminist theory. *Journalism & Mass Communication Quarterly* 83(2), 247–264 (2006)
- [17] Herring, S.: Gender differences in computer-mediated communication: Bringing familiar baggage to the new frontier. Retrieved April 29, 2002 (1994)
- [18] Jasso, G.: On gini's mean difference and gini's index of concentration. *American Sociological Review* pp. 867–870 (1979)
- [19] Johnson, T.J., Kaye, B.K., Bichard, S.L., Wong, W.J.: Every blog has its day: Politically-interested internet users perceptions of blog credibility. *Journal of Computer-Mediated Communication* 13(1), 100–122 (2007)
- [20] Karpowitz, C.F., Mendelberg, T., Shaker, L.: Gender inequality in deliberative participation. *American Political Science Review* 106(03), 533–547 (2012)
- [21] Marder, N.S.: Gender dynamics and jury deliberations. *Yale Law Journal* pp. 593–612 (1987)
- [22] Monnoyer-Smith, L.: The technological dimension of deliberation: a comparison between on and offline participation. *Connecting Democracy: Online Consultation and the Future of Democratic Discourse*, MIT Press, Cambridge, MA pp. 191–207 (2011)
- [23] Price, V.: Citizens deliberating online: Theory and some evidence. *Online deliberation: Design, research, and practice* pp. 37–58 (2009)
- [24] Sanders, L.M.: Against deliberation. *Political Theory* pp. 347–376 (1997)
- [25] Shane, P.: Turning gold into epg: Lessons from low-tech democratic experimentalism for electronic rulemaking and other ventures in cyberdemocracy. *Online deliberation: Design, research, and practice* pp. 149–162 (2009)
- [26] Siu, A.: Look who's talking: Examining social influence, opinion change, and argument quality in deliberation. ProQuest (2009)
- [27] Squires, J.: Deliberation, domination and decision-making. *Theoria: A Journal of Social and Political Theory* pp. 104–133 (2008)
- [28] Stromer-Galley, J., Wichowski, A.: Political discussion online. *The handbook of internet studies* 11, 168 (2011)
- [29] Trénel, M.: Facilitation and inclusive deliberation. *Online deliberation: Design, Research and Practice*, San Francisco: CSLI Publications (2009)
- [30] Wilson, P.A.: Deliberative planning for disaster recovery: Re-membering new orleans. *Journal of Public Deliberation* 5(1), 1 (2008)
- [31] Witschge, T.: Online deliberation: Possibilities of the internet for deliberative democracy. *Democracy online: The prospects for political renewal through the Internet* pp. 109–122 (2004)