Special Issue Article

Political Studies Association

POLITICAL STUDIES

Political Studies
1–23

DOI: 10.1177/0032321719890805
journals.sagepub.com/home/psx
SAGE

Demographics and (Equal?) Voice: Assessing Participation in Online Deliberative Sessions

Ryan Kennedy[1], Anand E Sokhey[2], Claire Abernathy[3], Kevin M Esterling[4], David MJ Lazer[5], Amy Lee[6], William Minozzi[7] and Michael A Neblo[7]

Abstract

Critics of deliberative democracy have worried that deliberation may mirror (or even exacerbate) inequalities in participation across categories such as gender, race, and age. Accordingly, we investigate the potential for technology and design to ameliorate these concerns, looking at the extent to which online deliberative sessions facilitate inclusive participation. In a large study of online deliberation (over 1600 participants nested in hundreds of online sessions), we examine differences in the amount and nature of participation across demographic categories, as well as the effect of forum characteristics on such differences. Though our results are mixed, we read them with cautious optimism: the online format is not immune to inequalities in participation and satisfaction, but we do not observe differences across some demographics, and most observed differences are substantively minor. Moreover, features of online deliberation environments show promise for addressing some of the problems plaguing in-person designs.

Keywords

communication technology, deliberation, participation, gender, race, age

Accepted: 18 October 2019

Introduction

Theorists have long argued that deliberation has the potential to improve democratic practice (e.g. Chambers, 1996; Habermas, 1984), and in recent years empirical research on

[1]Department of Political Science, University of Houston, Houston, TX, USA
[2]Department of Political Science, University of Colorado, Boulder, CO, USA
[3]Department of Political Science, Stockton University, Galloway, NJ, USA
[4]Department of Political Science, University of California, Riverside, CA, USA
[5]Department of Political Science, Northeastern University, Boston, MA, USA
[6]Institute for Democratic Engagement and Accountability, The Ohio State University, Columbus, OH, USA
[7]Department of Political Science, The Ohio State University, Columbus, OH, USA

Corresponding author:
Ryan Kennedy, Department of Political Science, University of Houston, 3551 Cullen Blvd., Room 447 PGH, Houston, TX 77204-3011, USA.
Email: rkennedy@uh.edu

deliberative practice has proliferated (for reviews, see Neblo, 2005, 2015; Thompson, 2008). Deliberation has been credited with producing a number of democratic goods, including increased political knowledge/learning (e.g. Barabas, 2004; Esterling et al., 2011), increased tolerance for opposing views (e.g. Gutmann and Thompson, 1996), and higher levels of faith in the democratic process (e.g. Fishkin, 1995; Neblo et al., 2018).

However, critics of the practice of deliberation have charged that traditionally politically underrepresented groups may not be heard or treated equally in group deliberative processes (e.g. Mansbridge, 1983; Sanders, 1997; Sunstein, 2006; Young, 2001). Most of these authors have focused squarely on how participation in deliberation sessions—in terms of attendance, quality of engagement, and satisfaction with the experience—suffers from similar biases seen in other types of political participation (Verba et al., 1997). Indeed, some recent empirical work has underscored these concerns: Karpowitz, Mendelberg, and Shaker find significant gender differences in deliberative participation in a series of small-group experiments (e.g. Karpowitz et al., 2012; Karpowitz and Mendelberg, 2014); gender composition and group decision rules interact to facilitate (diminish) female participation.

While these results are focused on one characteristic of participants (gender), they motivate a broader discussion of whether underrepresented groups differ in their in-deliberation participation. We deeply share these concerns, and in turn argue that such dysfunctions need not be a necessary part of deliberative practice. Rather, we stress the point that interactions (within small groups) are heavily conditioned on the design features that create the context for the interaction. Online platforms can be designed in a way that filters out cues regarding race, age or gender, and that prevent interruptions while others are speaking. Evidence to this effect can be found in the Neblo et al. (2018) findings of broad and equal participation in and satisfaction with carefully designed online town halls.

Leveraging 3 years of data gathered via the Common Ground for Action (CGA) online deliberation system (https://www.nifi.org/en/common-ground-action), we test whether the amount and nature of participation in online deliberative sessions varies across demographic groups, and whether these outcomes themselves depend on session-level characteristics such as the presence of a female group discussion moderator. Our study stands apart from many small-group studies not only in its (online) mode but also in its size and features: we analyze over 1600 participants nested in 275 unique deliberative sessions, each covering one of 20 topics; we include "moderate" facilitators and employ a decision rule that incorporates aspects of both unanimous rule and majority rule.

Although this study lacks explicit random assignment and experimental control, causal interpretations of our estimates may still be justified based on appropriate assumptions. We rely on naturally occurring variation in demographic composition of sessions, and participants lacked the ability to self-select into sessions based on these characteristics. That said, sessions also drew samples from different populations, and featured different topics that were likely to be differentially attractive to participants based on gender, race/ethnicity, or age (more generally, even things like the day of the week or the time of a session may have influenced demographic composition). Therefore, our analytic approach is to rely on random effects models with session-level intercepts, which are modeled with covariates that include recruitment method. Because these data are observational, causal inferences rely more heavily on analysis than they would in a randomized controlled trial. Nevertheless, these data provide the opportunity to focus on behavior in online deliberation at a scale that provides statistical power (and variance on substantive topic), all conducted through a carefully designed, state-of-the-art platform.

In the remainder of this article, we take several steps to gauge the nature of the deliberative experience in these online settings. First, we examine the frequency of participation in the sessions by gender, minority status, and age while considering the importance of two session-level factors: the proportion of women, and the presence of a female moderator. Second, we delve further into participant "voice," examining the extent to which patterns of speech differ across these demographic categories, as distinctive speech patterns might introduce biased responses in the text-based environment. Finally, we look at whether satisfaction with outcomes is conditioned by the demographic categories and the aforementioned session-level factors.

What emerges is a somewhat mixed set of findings with respect to participatory equality, though, on the whole, we find the results encouraging. Although we observe less participation among the non-white and the oldest participants, we find no significant net differences in total participation by gender, and that these non-differences between women and men are robust to session features such as group gender composition and moderator gender. We encourage scholars to continue to pursue the ideas we present, to more definitively link institutional features to outcomes. Importantly, we underscore that awareness of the shortcomings of the online environment does not condemn deliberation as a tool (Sunstein, 2006). Rather, documenting such dynamics provides guidance for deliberative practitioners; thinking through design problems can move the practice of online deliberation into better alignment with theoretical aspirations (Neblo et al., 2018).

The Promise and (Potential) Pitfalls of Online Deliberation

The Internet holds the potential to expand access to and involvement in democratic deliberation (Papacharissi, 2002; Stromer-Galley, 2017). With the capacity to directly engage a large number of citizens, Internet communication technologies can scale up deliberations by removing barriers to participation such as time and distance (Manosevitch, 2014; Smith et al., 2013; Strandberg, 2015). Indeed, online deliberations can reduce the resources needed to create deliberative opportunities compared to in-person sessions that physically bring people together (Min, 2007; Smith et al., 2013; Towne and Herbsleb, 2012). And, by extending the reach of deliberative opportunities, online deliberation may introduce participants to a more diverse range of views (Baek et al., 2011). In turn, this exposure to disagreement and differing opinions may meet important deliberative goals, challenging participants' assumptions and increasing their ability to articulate arguments in support of both their own views and opposing positions (Stromer-Galley, 2017; Stromer-Galley et al., 2015).

As with in-person events, the basic structure and organization of online forums can foster (or hinder) meaningful and inclusive deliberation. For example, the anonymity afforded to contributors in many online spaces may create the conditions for a wider range of views to be expressed, including opinions that may be unpopular or controversial (Ho and McLeod, 2008; Kim, 2006; Strandberg, 2015; Stromer-Galley, 2017; Stromer-Galley et al., 2015). That is, anonymity can be a structural feature of online deliberation that lowers barriers to entry for prospective deliberators who may otherwise feel judged for their expressions—this may have the benefit of facilitating more forthright discussion and encouraging broader participation (Rose and Sæbø, 2010; Stromer-Galley et al., 2015).

Of course, there are also reasons to be skeptical about online deliberation. For one, in practice online deliberation may involve relatively shallow consideration of political issues. In online discussions, users often present underdeveloped arguments in support of their comments (Ellis and Maoz, 2007; Hagemann, 2002; Weger and Aakhus, 2003), and some work has suggested that productive disagreement occurs less frequently in online settings compared to face-to-face deliberations (Stromer-Galley et al., 2015). Perhaps most notably, online forums have been criticized for their lack of civility, a concern directly connected to the ability of contributors to remain anonymous. While anonymity may free people to discuss unpopular or controversial opinions, by diminishing accountability it may also enable disrespectful or even vulgar exchanges that detract from deliberation (Coleman and Moss, 2012; Friess and Eilders, 2015; Rose and Sæbø, 2010; Stromer-Galley et al., 2015). But others argue that online deliberations do not inherently lack civility (Smith et al., 2013: 726), and some studies have found little evidence of such behavior (Neblo et al., 2018). Overall, then, there are reasons to suspect that the online environment may facilitate more reasoned and equal discussion, but there are also reasons to suspect that the online environment may present a different set of problems. Accordingly, we do not approach the subsequent analyses with directional expectations—we treat this study as an important exploratory effort, and an opportunity to reflect on the importance of deliberative design.

Features of Online Environments

The design of a deliberative space has consequences for its participants' engagement (e.g. Beauvais and Baechtiger, 2016; Karpowitz and Mendelberg, 2014; Strandberg, 2015). But how might the inclusion of certain design features in online forums affect the participation of different, often marginalized groups? More specifically, how might online deliberation facilitate more equal participation along lines of gender, race, and age? For one, text-based online environments remove visual cues of gender, age and race, and critically, disallow interruptions—one of the primary ways that the engagement of marginal groups has been undermined in in-person deliberation settings (Mendelberg et al., 2014; Rose and Sæbø, 2010). Even when first names are used as screen names, decades of research on information processing and memory (Asch and Ebenholtz, 1962; Curran and Doyle, 2011; McBride and Dosher, 2002), as well as some of our own research (see section A10 of the Supplementary Information), suggests that names on a screen are not nearly as strong as visual cues. Baek et al. (2011), for example, compared online and face-to-face deliberation, and found that online sessions tended to over-represent young male and white users. Despite this, online deliberators believed the online sessions were much more politically and racially diverse, and they perceived these sessions to have about equally (high) gender diversity. Similarly, interruption is more difficult in a text-based system, since comments are only entered upon the author's completion of their thought. This does not disallow someone from composing posts at the same time as someone else, but they cannot speak over or preempt a text post. Second, the aforementioned factor of anonymity may increase contributions from marginal group participants (Coleman and Moss, 2012; Price, 2006; Smith et al., 2013) by ameliorating issues involving social pressure.

However, other realities of the online experience also raise concerns about how effectively deliberative forums can facilitate free and fair inclusion of different perspectives. Gender hostility in online discussions can limit women's contributions—if speech/comments match gendered or other demographic stereotypes, an individual's contributions may be marginalized or sexualized (Harp and Tremayne, 2006; Stromer-Galley, 2017).

And, of course, anonymity has the potential to "cut both ways"—it can exacerbate the disrespectful or hostile tone in discussions, as contributors may feel more emboldened to express inflammatory views (Coleman and Moss, 2012; DiMaggio et al., 2001; Rose and Sæbø, 2010). In addition, in practice, online discussions could be dominated by a few vocal participants and fall short of realizing the ideal of broad participation across all contributors (Graham and Wright, 2014; Koop and Jansen, 2009).

Another significant challenge to the inclusivity of online deliberation is the digital divide (Strandberg, 2015). Internet access remains unequally distributed in the American population, with lower income, rural, and older populations in the United States less likely to have access to high speed Internet (DiMaggio et al., 2001; Vick, 2017), and older minorities being disproportionately affected (Stromer-Galley, 2017). Citizens without regular access to the Internet often lack technological skills to engage online (Olsson et al., 2003; Rose and Sæbø, 2010; Strandberg, 2015). These differences in access to and proficiency with Internet communication technologies have the potential to exacerbate the challenges that some underrepresented groups might already face within group interactions. That is, differential access and facility with online platforms may "lead to the emergence of new forms of inequality and social exclusion" (Smith et al., 2013: 711).

Evaluative Criteria

While the potential for online deliberation to increase the scalability of deliberation is clear (Neblo et al., 2018), less clear is its ability to address concerns about replicating biases present in other realms of political participation. To begin to take stock of online deliberation's ability to give equal presence and quality of experience to a diverse public, we (1) "listen" to what respondents had to say *in* the online sessions and (2) assess what they had to say *about* the online sessions. That is, we first examine whether women, non-whites, and older individuals commented at the same rates as young white males, and whether these patterns of online voice were conditioned by session-level characteristics (namely, the proportion of participants that are women, and whether the moderator was female). Next, we further scrutinize participation in the sessions by examining whether the text-based methods of our forums elicit demographically stereotypical discussion (through which bias could emerge), or whether they strip away linguistic markers. Finally, we consider whether women, older individuals, and underrepresented minorities found similar satisfaction with the deliberative outcomes (compared to young white males), and whether the same session-level characteristics conditioned these experiences.

Data: The Common Ground for Action Platform

As design considerations are fundamental to the deliberative possibilities of online spaces (Manosevitch, 2014; Stromer-Galley et al., 2015; Wright and Street, 2007), we turn to a description of the online platform that hosted the deliberative sessions featured here, CGA. The CGA platform was developed as a cooperative effort between academics from several institutions, the Kettering Foundation, and the National Issues Forum. We suspect that several design elements of this platform may help foster full and fair inclusion of marginal groups.

Each CGA session focuses on a single policy topic. Participants are provided with a non-partisan issue guide for the topic that takes participants through several policy options—with their advantages and tradeoffs—ahead of the forum. Within the session,

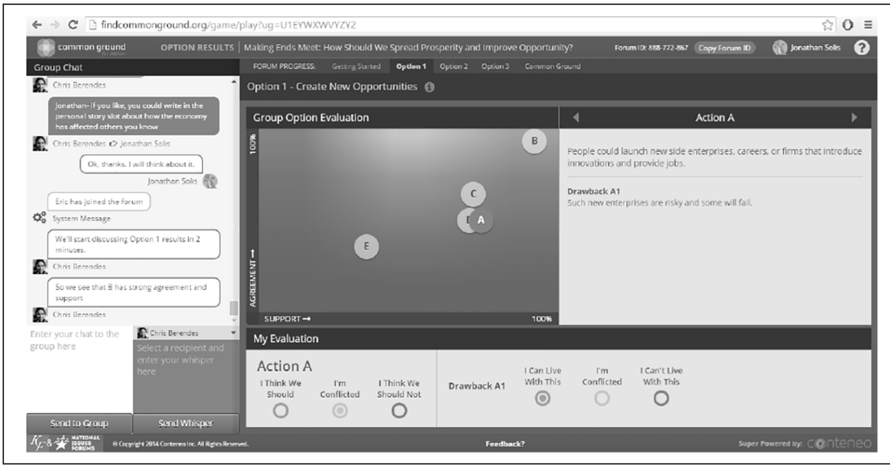

Figure 1. Example of Common Ground for Action Screen as Seen by Participants.
This screenshot shows an example of discussion on one option from a CGA on economic opportunity in the United States.

participants are asked whether they support each action, and whether they can live with a tradeoff. Preferences are revealed simultaneously and in aggregate. Participants then discuss each action with each other through a synchronous text chat platform. All sessions have a moderator who guides discussion and encourages participation. This moderator is the first to speak in the sessions and guides participants through the various topics of discussion. Participants can change their stance on an action at any time, and a graph tracks support in real time for the participants.

At the end of the session, those actions for which at least 75% of participants express support—and say they could tolerate the tradeoff—become the “common ground for action.” This decision rule places it somewhat between the majority and consensus rules of Karpowitz and Mendelberg (2014). In terms of mechanisms, however, both the numeric summaries and tendency for moderators to focus on areas of disagreement, suggest that these sessions are much closer to the majority rule sessions which “signal that conflict is acceptable” and are “based on a contest of interests” (Karpowitz and Mendelberg, 2014: 90). Figure 1 shows the presentation of the common ground to participants. Participants are then asked, on a scale from 1 to 5, how satisfied they would be if the common ground solution were to be implemented.[1]

Design of the CGA Platform

We suspect that several features of the CGA platform (and the user experience it provides) may work to expand participation, thereby meaningfully incorporating voices from marginalized groups:

- The text-based format for interaction removes visual cues of deliberator demographics and prevents interruptions. Users usually use their first names as the ID that appears on screen, meaning that some cues for gender and, to a lesser extent, race, and age remain. Users are also assigned avatars, but these are randomly assigned. However, as we noted in the "Data: The Common Ground for Action Platform" section, these cues are quite weak, and much less likely to affect substantive deliberation. Together, these may counter some of the basic barriers to equal participation by marginal groups (Mendelberg et al., 2014).

- All participants are provided factual reading material prior to the session; this may empower participation by helping individuals develop more reasoned opinions (e.g. Fishkin, 1995).
- Even before discussion of each option begins, each deliberator is asked to personally report their preferences on the options under consideration—thus, equal opportunities for participation across groups are built into the CGA platform. This task of indicating preferences ahead of discussion may allow CGA deliberators to more effectively explore areas of disagreement. In their study, comparing face-to-face and online deliberation, Stromer-Galley et al. (2015) note that obtaining baseline preferences "organically" in an online environment can be difficult.
- Moderators facilitate each CGA session and play an important role in guiding the conversation and soliciting opinions from participants. CGA facilitators involved in each session received training to fulfill the "moderate facilitator" role, as defined by Dillard (2013). The moderate facilitator acts as a "designated driver," moving the conversation along without adding new interpretations or editorial comments; moderators that follow this approach allow the group to more effectively learn about opposing views (Beauvais and Baechtiger, 2016; Dillard 2013). By facilitating the sessions, moderators are able to ensure that all participants have an equal chance to contribute, that different perspectives are expressed, and that no deliberator(s) dominates the conversation (Asenbaum, 2016; Beauvais and Baechtiger, 2016; Dillard, 2013).
- Synchronous chat allows real-time discussion on the topic. The design of the CGA setting alleviates the common concern that synchronous chat can lead to disorganized deliberations and low-quality argumentation.[2] The single-issue focus with conversation organized around structured policy options ameliorates concerns about lack of coherence in the sessions. Likewise, the active moderator can direct the conversation to ensure that different strands of an argument are considered, even if they arrive out of sequence.

Data Derived from CGA Sessions

The sessions we analyze took place across a range of populations and were originally conducted by different research groups for a variety of purposes over a period of 3 years. Many of the online forums took place with university students; some were open sessions where sessions were advertised to mailing lists of groups like Kettering and NIFI, a few took place with recruited stakeholders by NGOs around a particular issue, and a number of sessions were recruited from the general populations of Idaho, North Carolina, and Ohio with samples recruited by professional survey companies.[3] In the case of open and stakeholder sessions, participants self-selected into participation. For student sessions, participants were either offered extra credit, participated as part of a classroom exercise, or volunteered. The professional surveys tried to recruit a variety of participants (though they were not randomly selected) who were compensated for their participation with Amazon gift cards. Some participants in the open and student sessions participated in more than one CGA during this time. About 31% reported participating in more than one session, with about 11% participating in four or more.

While analyzing data from such a variety of sources poses some challenges, we attempt to address these issues in several ways detailed below. As we are analyzing these data after the fact, we do not have experimental control over assignment to session by demographic characteristics. Participants were either assigned to a group based on availability or the time of their arrival (in the case where more than one session was being conducted in the same time slot). Sessions ran between 40 and 200 minutes.[4] Pooling these online deliberations yields a total data set of 1609 participants nested in 275 sessions (with each session covering one of 20 topics, see section A3 of the Supplementary Information)—this is considerably larger and more varied than many empirical studies of deliberation in the literature.[5]

The forums provide us with several layers of information to analyze "online voice" (in-session behavior) and satisfactions with the experience. First, we have the discussion posts by the participants. This includes both the number and content of the posts (which are automatically saved from the session, and identified by a unique User ID and Session ID). These provide our first set of dependent variables: the number of chats posted by a participant and the number of words contributed to the session by the participant. Second, we have the content of the statements themselves. The words chosen by the participants make up the next part of our analysis. Finally, we have the expressed preferences of the participants and the eventual CGA developed in the session, along with the participants' expressed support for the common ground. We model satisfaction with the common ground, controlling for participants' expressed preferences.

Demographic variables were generally collected prior to the beginning of a session. In some cases, the demographic variables were collected in post-session surveys, which we acquired where we could.[6] Participants were asked to give their gender, age, and race. For gender, we create a dummy variable for analysis, where 1 indicates the participant is female and 0 indicates they are male. Age is an ordinal variable with four levels: under 30, 31–45, 4665, and over 65. We set the under 30 group as our baseline. For the race variable, participants were able to check one or more categories, or to input a category that was not mentioned. For ease of analysis, we converted this into a dummy variable where 1 indicates the participant is non-white and 0 indicates they are white.

In terms of demographics across these sessions, as might be expected, given that many of these sessions were conducted in social science classrooms on college campuses, the sample skews young, white, and female (see Figure A1 of the Supplementary Information). Participants under 30 made up about 75% of the sample, white participants made up about 64% of the sample, and women were about 65% of the participants. At the session level, the number of participants is constrained between 5 and 12, consistent with the definitions of small-group deliberation offered by previous scholars (e.g. Karpowitz and Mendelberg, 2014), and consistent with our own experience of where CGAs are effective. The median size for the sessions was 8. For the models that follow, we also included the proportion of participants in a session that were female. Moderator gender was encoded based on the username, and session moderators were split approximately equally between female, male, and those who did not specify their gender with their name (e.g. "CGA").

Because our data consist of individuals nested within deliberation sessions, we utilize multilevel models with a random effect for session, and session-level covariates (Gelman and Hill, 2006). This produces models of the form:

$$y_{ij} = \beta_{1j} + \beta_2(Female)_i + \beta_3(Age31-45)_i + \beta_4(Age46-64)_i$$
$$+\beta_5(65\,or\,older) + \beta_6(Non-white) + \beta_7(Preference\ deviation) + \kappa_s$$

where

$$\beta_{1j} = \gamma_{11} + \gamma_{12}(N\ participants)_j$$
$$+\gamma_{13}(Female\,moderator)_j + \gamma_{14}(Unspecified\ moderator)_j$$
$$+\gamma_{15}(Proportion\ female)_j + \gamma_{16}(Education)$$
$$+\gamma_{17}(Stakeholders) + \gamma_{18}(Survey) + \zeta_{1j}$$

At the individual level, $Female$ and $Non\text{-}white$ are indicator variables, and the age categories are indicator variables with those under 30 as the baseline. We also include a control for the individual's initial deviation from the average preferences of the group, since being in the minority might affect behavior (Myers, 2017). At the session level, $Female\,moderator$ and $Unspecified\ moderator$ are indicator variables for the indicated gender of the moderator; $Proportion\ female$ is the scaled proportion of women in the session; $N\ participants$ is the scaled number of participants in the session; $Education$ is an indicator for whether the forum was conducted using university students; $Stakeholders$ indicates if the forum involved community stakeholders in the issue; and $Survey$ indicates the forum's participants were recruited by a professional survey firm ($Open$ sessions are the baseline). And ζ_j is the random deviation of session j's mean measurement from the overall mean. By assumption $\zeta_j \sim N(0,\phi)$. For the models of satisfaction with the common ground, we add an additional variable that is the proportion of the common ground for which the individual participant gave approval, ranging from none of their approved actions being included (0) to all of them being included (1), in order to ensure that estimates are not simply picking up satisfaction based on pre-existing preferences. We also include a session-level variable for the correlation between pre-session preferences, since the level of disagreement has been found to affect the quality of deliberation (Esterling et al., 2015; Zhang, 2019). To aid in convergence and to allow for more meaningful visualization, we scale all the variables that are not already proportions or dummy variables by subtracting the mean and dividing by the standard deviation (Gelman and Hill, 2006).

Results

Online Voice: Rates of Participation

For rate of participation (Figure 2), we look at both the number of comments made by participants, and the total number of words they contributed to the session. By separating these two characteristics, we can note both differences in terms of likelihood of commenting, as well as differences in overall levels of participation. This is important, as these can point to different dynamics (with some participants providing a large number of relatively short comments, and others providing fewer, longer comments). The left-hand side of Figure 2 shows the relationship between demographic group and the number of comments offered during the session. We can see that women provide significantly fewer comments than men $(p < 0.0)$, with an average of 8% fewer posted comments

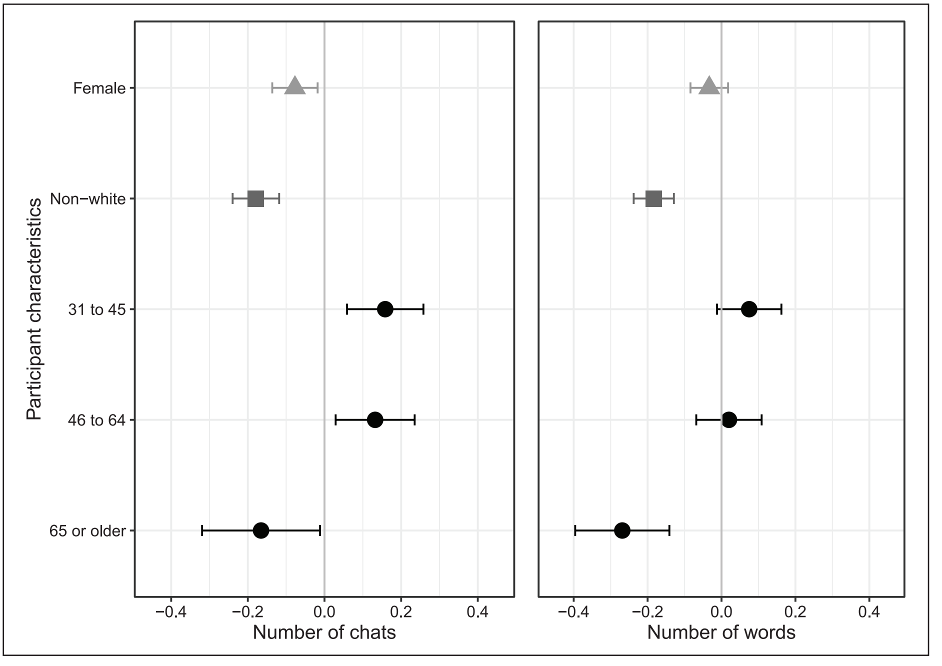

Figure 2. Effect of Demographic Characteristics on Scale of Participation.
Effects are the estimated effect of membership in each demographic characteristic on the scaled log of participation, with the 95% confidence intervals represented by the bars.

per person. The graph also shows that non-white participants also tend to provide fewer comments $(p < 0.01)$, with an average of 16% fewer posted comments. This is consistent across non-white racial categories (see section A9 of the Supplementary Information). Surprisingly, age seems to have a curvilinear effect. While the stereotype is that younger people are "digital native" and, therefore, more adaptive to technology, and the literature on other political activities suggests that older people participate more (Verba et al., 1995), we find that those in the 31–64 age group post about 14%–17% more chats in these sessions $(p < 0.05)$, with a steep falloff for those over 65 $(p < 0.01)$, who post about 15% fewer chats.

That said, looking at the number of words paints a slightly different picture. The right-hand plot of Figure 2 demonstrates that the significant differences in chats based on gender become insignificant, once we analyze the number of words posted in a session. Put differently, these results suggest that men and women differ more in terms of the type of participation than in the scale of participation, with women offering fewer, but longer posts. We observe a similar pattern among those under 31: the length of their posts brings them back in line with those in the 46–64 category in terms of overall participation (though they still lag somewhat relative to those in the 31–45 age range $(p < 0.1)$, by about 8%). The two groups that still show a lack of participation by this metric are non-white participants ($p < 0.01$, whose words typed in each session lag by about 17%) and those over 65 ($p < 0.01$, who post about 24% fewer words).

Figure 3 introduces session-level factors, testing whether women are more likely to participate when there are more women involved, when they observe other women in

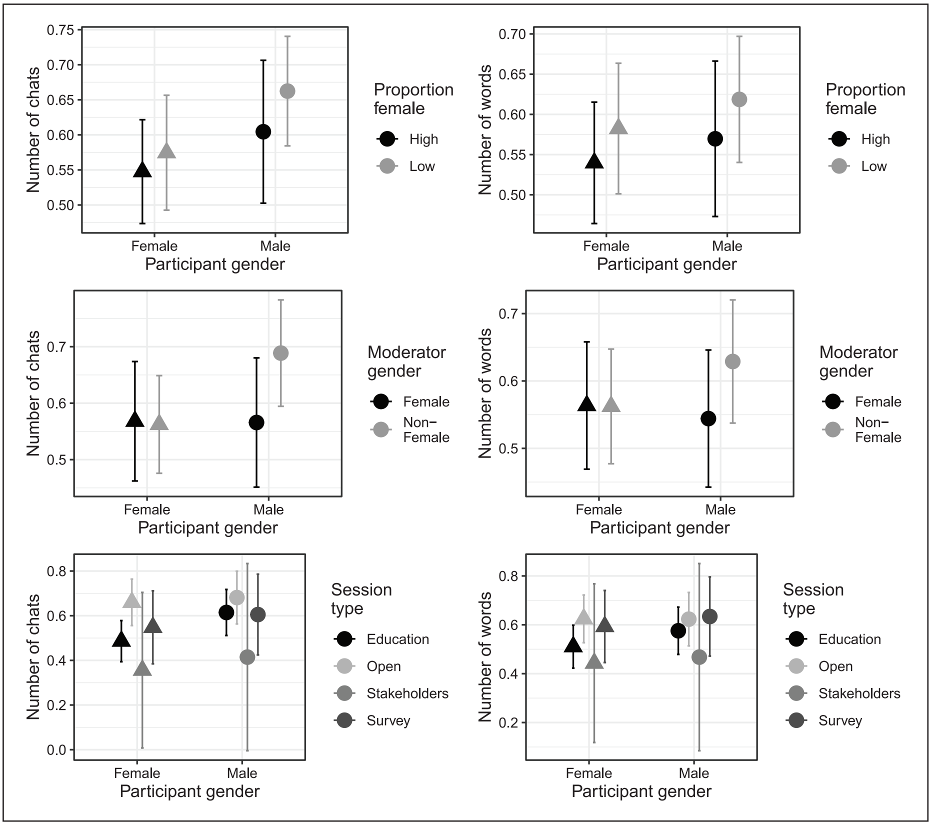

Figure 3. Effect of Session Characteristics on Female Participation.
The points denote the expected amount of participation (scaled and logged) when all other variables are held at their means, with the 95% confidence intervals represented by the bars.

moderator roles, or in certain types of sessions. While we are interested in a range of demographics, previous literature provides strong reasons to suspect that these factors will make a difference for gender participation (e.g. Karpowitz and Mendelberg, 2014). Although, empirical studies have produced mixed results, the literature on "critical mass theory" raises the possibility that women will participate more in sessions where they make up a higher proportion of the participants (Sarah and Mona, 2008). Similarly, there is some evidence to suggest that women in leadership roles empower other women to participate (Latu et al., 2013), raising the possibility that a female moderator will encourage more participation by women. The population from which participants are drawn may also influence the results, as sessions held in educational settings make it more likely for participants to know each other outside of the sessions. We evaluate possibilities by introducing cross-level interactions between participant gender and the proportion of participants who are female in the session, and between participant gender and whether moderator gender is identifiable as female.

Figure 3 displays differences in the expected number of words for each group at different values of the interaction variables. The top two charts show the differences in chats and words between men and women, conditioned on whether the proportion of women in the session is high (3rd quartile) or low (1st quartile). For both chats and words, there is no statistically significant difference in the rate of participation, dependent on the proportion of the session that is female. To the extent that we do notice a difference, both men and women seem to participate more when there are more women. In section A7 of the Supplementary Information, we test whether there is an effect of female majorities, finding no significant differences.[7] We also tested whether the number of non-white or over 65 participants changed participation patterns for these groups in section A8 of the Supplementary Information, and again found no significant interactions.

In looking at the other session-level characteristic, we do notice significant differences between female and non-female moderators, but not necessarily in a way that we might expect. That is, instead of observing that women participate more when the moderator can be identified as female, we find that *men participate less when the moderator is identified as female* ($p < 0.05$ for chats and $p < 0.10$ for words). In other words, we do observe that participation between genders is more equal with a female moderator, but the mechanism that produces this result appears to be the suppressing of male participation rather than the boosting of female participation. We also do not find a statistically significant difference for either number of chats or words depending on the population from which session participants were drawn.[8]

Overall, we do find some substantial differences in participation between demographic groups. Non-white and participants over 65 participate in the sessions at a substantially lower rate. Women have about the same volume of participation overall and a female moderator encourages gender parity, but, surprisingly, this seems to be through lowering male participation instead of through increasing female participation.

Online Voice: The Content of Participation

We next examine how much of the discursive content of these sessions was stereotypically gendered, or could be stereotypically classified by another demographic category. To do this, we look at the words that distinguish different groups in these sessions using two different approaches. The first is a dictionary-based approach, where pre-existing word categories are compared across groups. The second is a machine learning approach, support vector machines (SVM), that uses the words of the participant to predict the gender, race, and age of the participant.

For the dictionary-based approach, we leverage the dictionaries developed by Karpowitz and Mendelberg (2014) to measure the prevalence of "care" versus "financial" words; they found that women in their sessions used substantially more care words, while men used substantially more financial words. The care words dictionary is intended to capture discussion that relates to children, education, and health. The financial words dictionary is intended to capture discussion about money, wealth, and incentives. Both dictionaries stem from human coding about these categories and the words associated with them (see section A5 of the Supplementary Information). For the vast majority of session topics—poverty, the opioid crisis, immigration, criminal justice policy, and climate change—these dictionaries apply quite well. For some topics, like mass shootings, the dictionaries may not apply as well, but these are a very small minority of forums (see section A3 of the Supplementary Information).

Figure 4 presents the estimated relationship between demographic labels and the number of care and financial words used in participants' statements. Unlike Karpowitz and Mendelberg (2014), we do not find differences in terms of the use of financial words. We

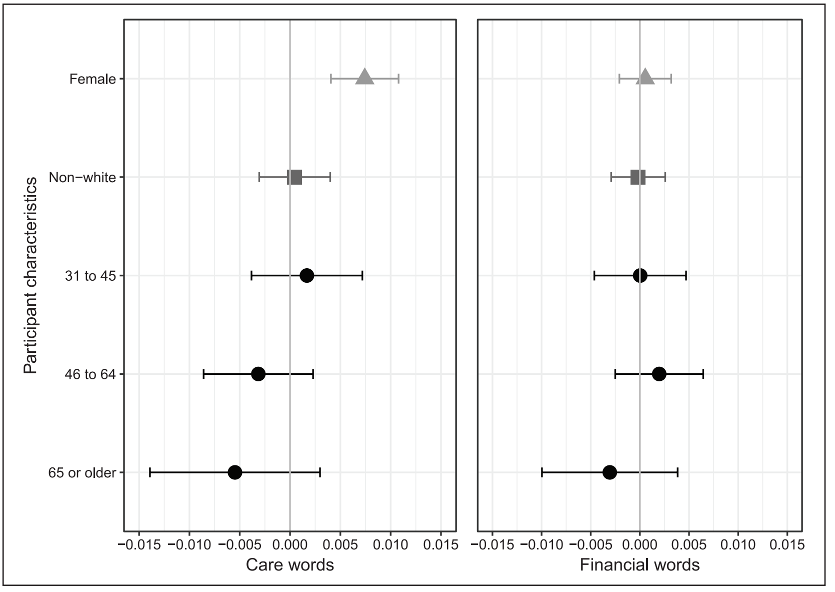

Figure 4. Estimated Proportion of Care and Financial Words Used in Statements.
Estimated coefficient and 95% confidence interval of proportion of words in participant statements that are in each dictionary.

do, however, find a statistically significant difference between men and women in terms of the number of care words. Of course, we would be careful not to overstate the substantive importance of this relationship: the findings suggest that women use about 0.8% more care words than men, which amounts to about 2.4 care words per session for an average participant (who uses 300 words in a session). While statistically significant—and suggesting some differences in terms of rhetorical focus—it is not entirely clear that this difference represents a substantively meaningful difference in rhetoric. None of the other demographic categories (race, age) produced a statistically significant difference in terms of these dictionaries.

The limitation to the dictionary-based approach is that it requires *a priori* defining of categories. A more flexible method is to use machine learning to determine which words are best at predicting the classification of participants, a method that might mimic participants' cognitive ability to classify comments by demographic category. We use SVM, which is a commonly used method in text classification. SVM is an extension of the support vector classifier, which maximizes the margin between observations from different classes within a chosen level of error—this error tolerance becomes a primary tuning parameter for the model. We utilize leave-one-out cross-validation to determine the ideal error tolerance. The sum of the errors in predicting the withheld observation produces the model's cross-validation error, which is minimized to set the tuning parameters.[9]

The results in Tables A11–A13 (in the Supplementary Information) show that the model does have some success in distinguishing between demographic categories. A naive model, which always predicts the modal category (female, young, and white) will be correct 64.4% of the time for gender, 65.1% of the time for age, and 55.5% of the time for race. The SVM model is correct about 66.3% of the time for gender (95% confidence interval = [64.0%, 68.6%]), 79.5% of the time for age (95% confidence interval = [77.3%, 81.6%]), and 67.6% of the time for race (95% confidence interval = [65.2%, 69.9%]). While not huge improvements, all but the gender differences are statistically significant at conventional levels, and the improvements in gender prediction are close to significant.

Looking at the words that are most important for classification in these models, we note some key differences, but the overall patterns of differences in discourse are not always clear (see Figure A4 of the Supplementary Information). In some cases, the differences between groups are very small—SVM is a highly non-linear statistical model, meaning it may capture some interactive effects that are difficult to capture as differences in proportions.

For gender, there are a few words that are particularly noteworthy. Women tend to state their agreement more, as indicated by their higher likelihood of using the word "agree." Female participants were also more likely to talk about drawbacks and about being conflicted. Few other differences are very distinguishable from the proportional usage. For example, men used the word "tax" slightly more, but women used the word "businesses" slightly more. To be clear, we are careful about placing too much emphasis on the substantive impact of these results. The clearest difference, in the use of "agree," amounts to 1 more use of the word agree on average for a female participant versus a male participant. The implication of this is also unclear, as the correlation between female participation and the amount of common ground generated is weak.[10]

For age, there are some interesting (if not completely consistent) differences. From a style perspective, younger participants are more likely to employ an informal style of communication, with greater prevalence of words like "yeah." Interestingly, they are more likely to emphasize "community" and "support." They also show a greater interest in "taxes." Older participants, by contrast, are more likely to emphasize "treatment" and "housing." As with gender, none of the direct proportions are all that telling, with the largest difference being the number of times younger people say "yeah."

Finally, for race, we note that non-white participants are more likely to talk about "people," "support," and "community." Like women, they are also more likely to "agree" during the sessions. However, as with gender we remain cautious about inferring too much substantive significance to these differences (they usually amount to only one additional use per session at most).

In sum, we are able to note differences between groups in the language they use in the sessions, but the substantive implications of these differences appear to be very small—usually only one additional use per session.

Satisfaction with Outcomes

Finally, we turn our attention from what participants did in the deliberative sessions, to what they had to say about the ultimate outcomes of the CGAs. Figure 5 looks at how participants rated the final "common ground" produced in the forums. It shows that there are few meaningful differences between groups in terms of their satisfaction with the deliberative outcomes. Women and older participants are, on average, more satisfied, but these differences are not statistically significant. Non-white participants are the only group to show a significant difference in their level of satisfaction ($p < 0.05$), with them reporting significantly less satisfaction with outcomes relative to their white counterparts.

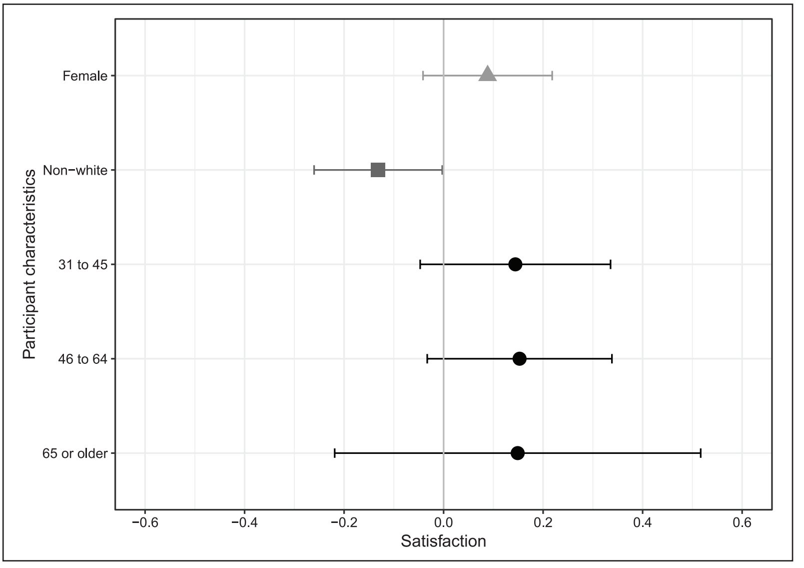

Figure 5. Effect of Demographics on Satisfaction with Common Ground.
The points are the estimated effect of membership in each demographic characteristic on the satisfaction score given for the common ground created in the session, with the 95% confidence intervals represented by the bars.

Figure 6 graphs the interactions between gender and session-level characteristics (the proportion of female participants and whether the moderator was female; we also look at the level of female participation (words from female participants divided by total words)). The results suggest that these session-level variables can change the way that the genders perceive the experience. On the one hand, a female moderator appears to increase the satisfaction women report with the final outcome, although this effect does not quite reach standard levels of statistical significance. On the other hand, the amount of female participation does significantly impact the level of satisfaction expressed by the different genders ($p < 0.05$), but, again, in a surprising manner. Greater female participation does not appear to impact women's satisfaction with the common ground outcome, but again exerts a leveling effect by significantly reducing male satisfaction.

In terms of satisfaction with the outcomes of the session, these results suggest that non-white participants are generally less satisfied with the outcomes. Women are somewhat more satisfied when there is a female moderator, and, discouragingly, men are less satisfied as female participation increases.

Discussion

Looking across the three analyses, we see an intriguing (and mixed) set of results. In our analysis of participation in the CGAs, we did not replicate all the gender-based inequalities observed in some previous studies (e.g. Karpowitz et al., 2012; Karpowitz and Mendelberg, 2014). Instead, we noted that female participants made fewer statements, but that their statements were longer, such that their total participation roughly equaled that of male participants. Although we cannot be certain, it would seem likely that this is a function of the text-based interface, which does not allow for interruption (a key factor

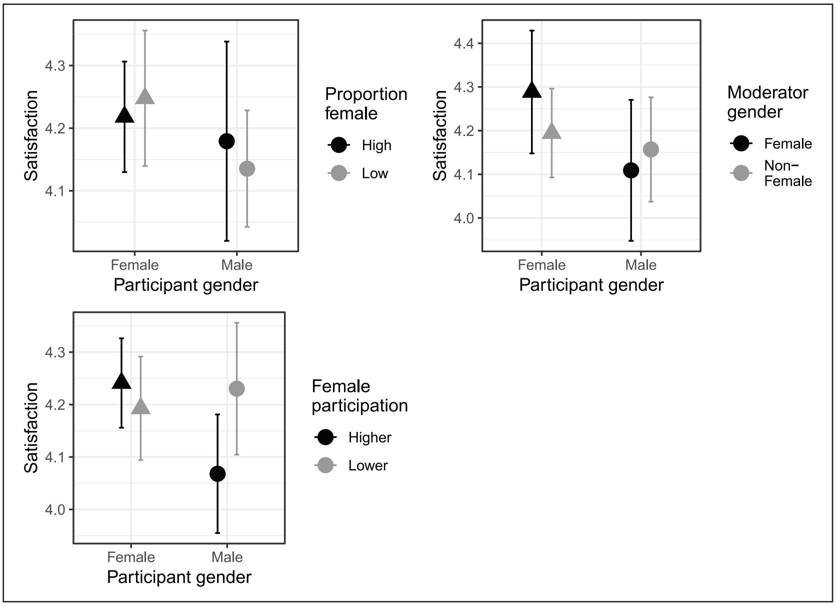

Figure 6. Interaction Between Demographics and Session Characteristics on Satisfaction with Common Ground.
In these charts, the points denote the expected satisfaction when all other variables are held at their means, with the 95% confidence intervals represented by the bars.

in explaining lower female participation in in-person events). While the CGA decision rules do not perfectly replicate those of previous studies, the numeric summaries on individual issues and clear signal that disagreement is acceptable suggest that this result is not due to a view that inclusion and consensus is required (Karpowitz and Mendelberg, 2014). At the session level, we observed an interesting dynamic whereby female moderators seem to introduce gender-equality by suppressing male participation. While we could not conduct a similar analysis by the race of the moderator, both because a number of moderators used non-specific aliases and, at least for the sessions where we know the moderator's race, the moderators were predominately white, future research should make an effort to evaluate the effect of recruiting more racially diverse moderators.

On a more problematic note, we did see inequalities emerge based on age and race. Like the gender dynamics, we noted qualitative differences among younger respondents compared to other age groups—they offered fewer but longer comments than those 31–64. Those over the age of 65 also offered fewer chats, and their chats tended to be shorter than those in all other age groups. This supports the hypothesis that age-related biases run counter to traditional forms of political participation like voting and campaign contributions, where older Americans are more likely to actively participate (Verba et al., 1995). Likewise, we did find persistent differences between white and non-white participants, with non-white participants participating less by all metrics analyzed.

With respect to the content of communication, we did find some differences that fit previous literature and conventional wisdom. For example, female participants were more likely to use care words in the sessions, but they were no less likely to use financial words. That said, the whole of the evidence for demographic differences based on what was said in the CGAs is fairly weak. Substantive differences between men and women were small, and differences in more general word usage based on other demographic characteristics appeared small, and only emerged after the application of a very non-linear model. Thus, while we do find some differences that fit with traditional demographic stereotypes, our results seem more in line with previous works highlighting the context-dependency of language—observed differences are not always meaningful (e.g. Bradley, 1981; Weatherall, 2002). The results from our machine learning algorithms and linguistic dictionaries seem to back up the conclusions of Newman et al. (2008: 229–230) on gendered language, who reviewed a range of studies and approaches to gender classification: the differences in language usage are "in the range generally considered small" and "[t]he fact that we are confronted with these differences every day yet fail to notice them highlights the degree to which they are part of everyday life." In analyzing transcripts we do find some evidence of differences, but we would be hard pressed to call these substantively meaningful. With this said, our work is just a starting place, with a number of methods for detecting specific differences in use of fuzzy qualifiers, political sophistication (Benoit et al., 2019), and complex thinking (Erisen et al., 2018) that might be usefully applied in the study of political demographic differences. Findings may also differ between formal and informal settings, a pattern already noticed in computational linguistics (Newman et al., 2008).

Finally, there are some intriguing differences in satisfaction with session outcomes. Non-white participants seem to be less satisfied with the overall outcomes than their white counterparts, while male participants seem to be less satisfied with the outcome when women participate at higher rates. For now, we can only speculate as to why these differences emerge—they suggest important and potentially fruitful avenues for future research.

Conclusion

Our analysis of CGA sessions represents, by an order of magnitude, the largest analysis of online deliberative behavior yet conducted, and identifies some differences in "voice" and experience across demographic characteristics. From a normative perspective, the results contain good and bad news, for while gender differences were mostly muted, other consistencies emerged between minority treatment in conventional political participation (e.g. Verba et al., 1995) and online deliberation. While additional studies are needed before we should draw definitive lessons, the results do underscore the fact that designers of online systems cannot assume equality of experience by taking people out of the room and putting them in an online environment (even with a state-of-the-art platform that is set to minimize demographic cues and social pressure, eliminate interruption, introduce organization, and assist with moderator duties). Moreover, the results hint at some additional factors, such as men's response to greater female participation that must be addressed within a broader social context.

While our analysis points to several challenges, it is a far cry from suggesting—as some scholars have (e.g. Sunstein, 2006)—that the deliberation project is futile. Indeed, we view the glass as "half-full." Online deliberation appears to have just as many benefits as drawbacks. Substantively, our findings can be taken to mean that, to the extent that there are differences in in-person deliberation, online deliberation may help to erase some of those differences. And, it is likely that with further design enhancements, broadened access to online participation, and increasing public familiarity with various interfaces, online deliberation could reduce societal inequities further. Moreover, as with other forms of participation, awareness of problems is key to explicitly addressing them (Neblo et al., 2018). In the end, this study provides a starting place for further exploration into the design and deployment of online deliberation platforms. We urge practitioners of deliberation to take these findings and use them to address these (and other) issues.

Future research should look into the impacts of providing versus withholding demographic cues, the dynamics of online text versus online speech, and incorporate a broader range of attitudinal and demographic characteristics in their data collection and analysis. Online deliberation also opens opportunities for the strategic use of random assignment to explore these options and boost internal validity. While not a panacea, the opportunities afforded by online deliberation are manifold. The Internet is now inextricably linked to the democratic project, and the latter will be strengthened by designing more effective ways to bring people together online.

Acknowledgements

The authors would like to thank the Kettering Foundation and the National Issues Forums for their assistance with data collection; Jon Green, Jon Kingzette, Myriam Shiran, Danielle Niangar, Jonathan A. Solis, Justin Freebourn, Laura Moses, Matthew Ward, Rogette Saba, Jack Brown, Philip Waggoner, Stephen Post, Rogette Saba, and a number of others for their research assistance; and John Gastil, workshop participants at Northeastern University, and two anonymous reviewers for their comments. All errors are the authors'.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded in part by Humility & Conviction in Public Life, a project of the University of Connecticut sponsored by the John Templeton Foundation. This research is also based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via 2017-17061500006 and 2017-17071900005. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the US Government. The US Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

ORCID iD

Ryan Kennedy https://orcid.org/0000-0003-1881-6887

Supplementary Information

Additional supplementary information may be found with the online version of this article.

Contents

A3. Topics of Sessions

Figure A3. Proportion of Sessions Conducted Under Each Topic.

A4. Full Results

Table A2. The Effect of Participant Characteristics on the log Number of Chats and Words in CGA Session.

Table A3. The Effect of Participant Characteristics on Satisfaction with Common Ground.

A5. Care and Finance Dictionaries.

Table A4. The Effect of Participant Characteristics on the Use of Care and Financial Words.

A6. Methods Description and Additional Results: Content of Sessions and Chats

Table A5. Confusion Matrix for Completely Naive Model of Gender.

Table A6. Confusion Matrix for Completely Naive Model of Age.

Table A7. Confusion Matrix for Completely Naive Model of Race.

Table A8. Confusion Matrix for Naive Bayes Model of Gender.

Table A9. Confusion Matrix for Naive Bayes Model of Age.

Table A10. Confusion Matrix for Naive Bayes Model of Race.

Table A11. Confusion Matrix for SVM Model of Gender.

Table A12. Confusion Matrix for SVM Model of Age.

Table A13. Confusion Matrix for SVM Model of Race.

Figure A4. Proportion of Words in Top-25 for Classification by Demographic Characteristics.

A7. Effect of Female Majority

Table A14. Effect of Female Majorities in Session on Female Participation.

A8. Effect of Proportion of Non-white Participants and Over 65 Participants

Table A15. Testing Interaction with Proportion Non-white and Proportion over 65.

A9. Disaggregation of Racial Categories

Table A16. Testing Disaggregation of Racial Categories.

A10. Identification of Textual Gender Information

A11. Results Split by Educational Setting Sessions and Other Sessions

Table A17. The Effect of Participant Characteristics on the log Number of Chats and Words in CGA Sessions in Education Setting only.

Table A18. The Effect of Participant Characteristics on the log Number of Chats and Words in CGA Sessions in Non-Education Setting only.

Notes

1. A more comprehensive discussion of the design of the CGA platform can be found in Gastil (2016). For a video of how this works, the interested reader can go to https://www.nifi.org/en/common-ground-action.
2. Stromer-Galley et al. (2015) and Weger and Aakhus (2003) argue that synchronous chat, where participants can simultaneously comment and their contributions are added to a running transcript of the conversation, may limit the coherence of discussion and make it more difficult to explore areas of disagreement. In this format, comments on any given topic or question may appear several lines apart, requiring participants to "sift through several lines of dialogue in order to find messages . . . relevant to their discussion," and making it difficult to sustain focus on particular points of interest or disagreement (Weger and Aakhus, 2003).

3. See section A2 of the Supplementary Information for details of participant profiles and research purposes. The professional samples were recruited by Qualtrics and Naviscent.
4. Some sessions ran longer or shorter than these thresholds, but were excluded because they were quite rare and had unusual characteristics (e.g. they fell below the threshold of participants set by the researcher and were cut short or there were technical problems).
5. Indeed, much of the evidence regarding disparities in group participation has come from work looking at a relatively small number of sessions, and often covering a single topic. For example, Karpowitz et al. (2012) had 470 student participants in 94 sessions discussing policies for redistribution, with some analyses using less than 20 sessions. Neblo et al. (2018) had 394 participants in 20 online sessions with Members of Congress (discussing immigration policy). To be clear, we do not mention these comparisons to be critical of these authors' decisions and designs (which were made for a variety of reasons, and offer different contributions). Rather, we would simply point out that having a smaller number of sessions in a given study poses power problems for estimating the impact of demographic characteristics, especially when testing whether patterns of participation vary by session-level characteristics.
6. These post-session surveys also collected other demographic variables, but they only covered 22% of the participants. We also explored inferring gender, race and age using participants' screen names, but did not incorporate this into our analysis, since doing so would introduce substantial noise with little increase in sample size.
7. Unlike Karpowitz and Mendelberg (2014), who analyze experimental evidence, we exploit observed variation in gender mix across CGA sessions. Thus, we do not claim to attempt a direct replication of their study. That said, there was considerable variation in the mix of genders by session, with more than 30 sessions populated exclusively by women and almost 50 in which women were in the minority.
8. In section A11 of the Supplementary Information, we split the population between educational and non-educational settings. The results for gender are substantively consistent across groups, and the finding on race varies only in terms of magnitude. The age finding does differ, which is not surprising, given that three-fourth of our over 65 participants are from non-educational settings. The observed effect for 31–45-year-olds is primarily present in the educational groups, while the over 65 results are primarily driven by the non-education setting sessions.
9. This approach is the same as that recommended by Hainmueller and Hazlett (2014). We utilized the automated tuning algorithm from the *caret* package in R to set the tuning parameters based on leave-one-out errors (Kuhn, 2008). A more comprehensive explanation of these methods is available in section A6 of the Supplementary Information. In our application, SVM outperforms a common alternative learner (naive Bayes, see section A6 of the Supplementary Information).
10. The correlation between the percentage of participants who are female and the percentage of actions in the final common ground is $r = -0.089$, while the correlation for the percentage of total words that come from female participants is $r = 0.0007$.

References

Asch SE and Ebenholtz SM (1962) The Principle of Associative Symmetry. *Proceedings of the American Philosophical Society* 106 (2): 135–163.

Asenbaum H (2016) Facilitating Inclusion: Austrian Wisdom Councils as Democratic Innovation between Consensus and Diversity. *Journal of Public Deliberation* 12 (2): 7.

Baek YM, Wojcieszak M and Carpini MXD (2011) Online Versus Face-to-Deliberation: Who? Why? What? with What Effects? *New Media & Society* 14 (3): 363–383.

Barabas J (2004) How Deliberation Affects Policy Opinions. *American Political Science Review* 98 (4): 687–701.

Beauvais E and Baechtiger A (2016) Taking the Goals of Deliberation Seriously: A Differentiated View on Equality and Equity in Deliberative Designs and Processes. *Journal of Public Deliberation* 12 (2): 2.

Benoit K, Munger K and Spirling A (2019) Measuring and Explaining Political Sophistication through Textual Complexity. *American Journal of Political Science* 63 (2): 491–508.

Bradley PH (1981) The Folk-Linguistics of Women's Speech: An Empirical Examination. *Communication Monographs* 48 (1): 73–90.

Chambers S (1996) *Reasonable Democracy*. Ithaca, NY: Cornell University Press.

Coleman S and Moss G (2012) Under Construction: The Field of Online Deliberation Research. *Journal of Information Technology & Politics* 9 (1): 1–15.

Curran T and Doyle J (2011) Picture Superiority Doubly Dissociates the ERP Correlates of Recollection and Familiarity. *Journal of Cognitive Neuroscience* 23 (5): 1247–1262.

Dillard KN (2013) Envisioning the Role of Facilitation in Public Deliberation. *Journal of Applied Communication Research* 41 (3): 217–235.

DiMaggio P, Hargittai E, Neuman WR, et al. (2001) Social Implications of the Internet. *Annual Review of Sociology* 27 (1): 307–336.

Ellis DG and Maoz I (2007) Online Argument between Israeli Jews and Palestinians. *Human Communication Research* 33 (3): 291–309.

Erisen C, Redlawsk DP and Erisen E (2018) Complex Thinking as a Result of Incongruent Information Exposure. *American Politics Research* 46 (2): 217–245.

Esterling K, Neblo M and Lazer D (2011) Means, Motive, and Opportunity in Becoming Informed about Politics: A Deliberative Field Experiment with Members of Congress and Their Constituents. *Public Opinion Quarterly* 75: 483–503.

Esterling KM, Fung A and Lee T (2015) How Much Disagreement Is Good for Democratic Deliberation? *Political Communication* 32 (4): 529–551.

Fishkin JS (1995) *The Voice of the People*. New Haven, CT: Yale University Press.

Friess D and Eilders C (2015) A Systematic Review of Online Deliberation Research. *Policy & Internet* 7 (3): 319–339.

Gastil J (2016) Building a Democracy Machine: Toward an Integrated and Empowered Form of Civic Engagement. Available at: http://ash.harvard.edu/files/ash/files/democracy_machine.pdf (accessed 21 October 2018)

Gelman A and Hill J (2006) Applied Regression and Multilevel/hierarchical Models. Cambridge: Cambridge University Press.

Graham T and Wright S (2014) Discursive Equality and Everyday Talk Online: The Impact of "superparticipants.*" Journal of Computer Mediated Communication* 19 (3): 625–642.

Gutmann AG and Thompson DF (1996) *Democracy and Disagreement: Why Moral Conflict Cannot Be Avoided in Politics.* Cambridge, MA: Harvard University Press.

Habermas J (1984) *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society.* Cambridge, MA: MIT Press.

Hagemann C (2002) Participation in and Contents of Two Dutch Political Party Discussion Lists on the Internet. *Javnost—the Public* 9 (2): 61–76.

Hainmueller J and Hazlett C (2014) Kernel Regularized Least Squares: Reducing Misspecification Bias with a Flexible and Interpretable Machine Learning Approach. *Political Analysis* 22 (2): 143–168.

Harp D and Tremayne M (2006) The Gendered Blogosphere: Examining Inequality Using Network and Feminist Theory. *Journalism and Mass Communication Quarterly* 83 (2): 247–264.

Ho SS and McLeod DM (2008) Social-Psychological Influences on Opinion Expression in Face-To and Computer-Mediated Communication. *Communication Research* 35 (2): 190–207.

Karpowitz C, Mendelberg T and Shaker L (2012) Gender Inequality in Deliberative Participation. *American Political Science Review* 106: 533–547.

Karpowitz CF and Mendelberg T (2014) *The Silent Sex: Gender, Deliberation, and Institutions.* Princeton, NJ: Princeton University Press.

Kim JY (2006) The Impact of Internet Use Patterns on Political Engagement: A Focus on Online Deliberation and Virtual Social Capital. *Information Polity* 11 (1): 35–49.

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

Kuhn M (2008) Building Predictive Models in R Using the Caret Package. *Journal of Statistical Software* 28 (5): 1–26.

Latu IM, Mast MS, Lammers J, et al. (2013) Successful Female Leaders Empower Women's Behavior in Leadership Tasks. *Journal of Experimental Social Psychology* 49 (3): 444–448.

McBride DM and Dosher BA (2002) A Comparison of Conscious and Automatic Memory Processes for Picture and Word Stimuli: A Process Dissociation Analysis. *Consciousness and Cognition* 11 (3): 423–460.

Manosevitch I (2014) The Design of Online Deliberation: Implications for Practice, Theory and Democratic Citizenship. *Journal of Public Deliberation* 10 (1): 9.

Mansbridge JJ (1983) *Beyond Adversary Democracy.* Chicago, IL: University of Chicago Press.

Mendelberg T, Karpowitz CF and Oliphant JB (2014) Gender Inequality in Deliberation: Unpacking the Black Box of Interaction. *Perspectives on Politics* 12 (1): 18–44.

Min SJ (2007) Online Vs. Face-To Deliberation: Effects on Civic Engagement. *Journal of Computer-mediated Communication* 12 (4): 1369–1387.

Myers CD (2017) Interests, Information, and Minority Influence in Deliberation. *The Journal of Politics* 79 (3): 804–822.

Neblo MA (2005) Thinking through Democracy: Between the Theory and Practice of Deliberative Politics. *Acta Politica* 40 (2): 169–181.
Neblo MA (2015) *Deliberative Democracy between Theory and Practice*. Cambridge, MA: Cambridge University Press.
Neblo MA, Esterling KM and Lazer DM (2018) *Politics with the People: Building a Directly Representative Democracy.* New York: Cambridge University Press.
Newman ML, Groom CJ, Handelman LD, et al. (2008) Gender Differences in Language Use: An Analysis of 14,000 Text Samples. *Discourse Processes* 45: 211–236.
Olsson T, Sandström H and Dahlgren P (2003) An Information Society for Everyone? *Gazette* 65 (4–5): 347–363.
Papacharissi Z (2002) The Virtual Sphere: The Internet as a Public Sphere. *New Media & Society* 4 (1): 9–27.
Price V (2006) Citizens Deliberating Online: Theory and Some Evidence. In: Davies T and Gangadharan SP (Eds) Online Deliberation: Design, Research and Practice. Stanford, CA: CSLI Publications, pp.1–22.
Rose J and Sæbø Ø (2010) Designing Deliberation Systems. *The Information Society* 26 (3): 228–240.
Sanders LM (1997) Against Deliberation. *Political Theory* 25 (3): 347–376.
Sarah C and Mona LK (2008) Critical Mass Theory and Women's Political Representation. *Political Studies* 56 (3): 725–736.
Smith G, John P and Sturgis P (2013) Taking Political Engagement Online: An Experimental Analysis of Asynchronous Discussion Forums. *Political Studies* 61 (4): 709–730.
Strandberg K (2015) Designing for Democracy? An Experimental Study Comparing the Outcomes of Citizen Discussions in Online Forums with Those of Online Discussions in a Forum Designed According to Deliberative Principles. *European Political Science Review* 7 (3): 451–474.
Stromer-Galley J (2017) Political Discussion and Deliberation Online. In: Kenski K and Jamieson KH (eds) *The Oxford Handbook of Political Communication*. Oxford: Oxford University Press, pp.837–850.
Stromer-Galley J, Bryant L and Bimber B (2015) Context and Medium Matter: Expressing Disagreements Online and Face-To in Political Deliberations. *Journal of Public Deliberation* 11 (1): 1.
Sunstein CR (2006) *Infotopia: How Many Minds Produce Knowledge*. Oxford: Oxford University Press.
Thompson D (2008) Deliberative Democratic Theory and Empirical Political Science. *Annual Reviews of Political Science* 11: 497–520.
Towne WB and Herbsleb JD (2012) Design Considerations for Online Deliberation Systems. *Journal of Information Technology & Politics* 9 (1): 97–115.
Verba S, Brady H and Schlozman K (1995) *Voice and Equality: Civic Volunteerism in American Politics.* Cambridge, MA: Harvard University Press.
Verba S, Burns N and Schlozman KL (1997) Knowing and Caring about Politics: Gender and Political Engagement. *Journal of Politics* 59 (4): 1051–1072.
Vick K (2017) The Digital Divide: A Quarter of the Nation Is without Broadband. Time, 30 March. Available at: https://time.com/4718032/the-digital-divide/
Weatherall A (2002) Gender, Language, and Discourse. London: Routledge.
Weger H and Aakhus M (2003) Arguing in Internet Chat Rooms: Argumentative Adaptations to Chat Room Design and Some Consequences for Public Deliberation at a Distance. *Argumentation and Advocacy* 40 (1): 23–38.
Wright S and Street J (2007) Democracy, Deliberation and Design: the Case of Online Discussion Forums. *New Media & Society* 9 (5): 849–869.
Young I (2001) Activist Challenges to Deliberative Democracy. *Political Theory* 29 (5): 670–690.
Zhang K (2019) Encountering Dissimilar Views in Deliberation: Political Knowledge, Attitude Strength, and Opinion Change. *Political Psychology* 40 (2): 315–333.

Author Biographies

Ryan Kennedy is Associate Professor of Political Science at the University of Houston and the Director of the Machine-Assisted Human Decision-making (MAHD) Lab. His research focuses on technology in politics and decision-making, machine learning, and computational social science.

Anand E Sokhey is Associate Professor of Political Science at the University of Colorado at Boulder and the Director of the Leroy Keller Center for the Study of the 1st Amendment. His research focuses on how social and contextual factors affect vote choice, political participation, and opinion formation in mass publics.

Claire Abernathy is an Assistant Professor of Political Science at Stockton University. Her research focuses on how members of Congress understand and respond to constituency opinion.

Kevin M Esterling is Professor of Public Policy and Political Science, and the Director of the Center for Technology, Communication and Democracy (TeCD), at the University of California, Riverside. His research focuses on technology and institutional design for communication in democratic politics.

David MJ Lazer is University Distinguished Professor of Political Science and Computer and Information Science, Northeastern University, and Co-Director for the NULab for Texts, Maps, and Networks. His research focuses on the nexus of network science, computational social science, and collaborative intelligence.

Amy Lee is Associate Director of the Institute for Democratic Engagement and Accountability at The Ohio State University.

William Minozzi is an Associate Professor of Political Science at The Ohio State University. His principal research interests are in legislative politics, deliberative democracy, political communication, and formal theory.

Michael A Neblo is Professor of Political Science and (by courtesy) Philosophy, Communication, and Public Policy at The Ohio State University where he directs the Institute for Democratic Engagement and Accountability (IDEA). His research focuses on deliberative democracy and political psychology.