

Effects of Comment Curation and Opposition on Coherence in Online Policy Discussion

Brian McInnis¹, Dan Cosley¹, Eric Baumer², Gilly Leshed¹

¹Information Science, Cornell University

²Computer Science & Engineering, Lehigh University

¹{bjm277, drc44, gl87}@cornell.edu, ²ericpsb@lehigh.edu

ABSTRACT

Public concern related to a policy may span a range of topics. As a result, policy discussions struggle to deeply examine any one topic before moving to the next. In policy deliberation research, this is referred to as a problem of topical *coherence*. In an experiment, we curated the comments in a policy discussion to prioritize arguments *for* or *against* a policy proposal, and examined how this curation and participants' initial positions of *support* or *opposition* to the policy affected the coherence of their contributions to existing topics. We found an asymmetric interaction between participants' initial positions and comment curation: participants with different initial positions had unequal reactions to curation that foregrounded comments with which they disagreed. This asymmetry implies that the factors underlying coherence are more nuanced than prioritizing participants' agreement or disagreement. We discuss how this finding relates to curating for coherent disagreement, and for curation more generally in deliberative processes.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous.

Author Keywords

deliberation; curation; preferences; coherence; agreement; openness

INTRODUCTION

Coherence in online policy discussion refers to the *consistency of the topics within a thread of comments* [67]. When discussion participants are regularly off-topic, or move away from topics too quickly, the behavior leads to an incoherent discussion that cannot deeply consider a policy issue. Too much attention on a single topic is also limiting, although in practice online policy discussions often result in far more than one topic being deeply considered [12, 14, 28, 31, 43, 71].

ACM acknowledges that this contribution was authored or co-authored by an employee, or contractor of the national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only. Permission to make digital or hard copies for personal or classroom use is granted. Copies must bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. To copy otherwise, distribute, republish, or post, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

GROUP '18, January 7–10, 2018, Sanibel Island, FL, USA

© 2018 ACM. ISBN 978-1-4503-5562-9/18/01... 15.00

DOI: <https://doi.org/10.1145/3148330.3148348>

Low coherence in online discussions about political issues has been observed across various forms of digital media: newspaper comment threads [14, 43], political [12] and non-political discussion forums [28], social platforms [31, 44], and field trials of advanced policy deliberation systems [29]. Although there are many recommendations [3, 11, 13, 70, 73] and prototype designs to encourage people using an online discussion system to build on the existing discussion [35, 38, 75], this empirical research indicates that *supporting coherence remains an unresolved design challenge*.

Here we examine the relationship between coherence and a commonly referenced design lever to affect the ways that people contribute to an online discussion: *content curation* [1]. Curating a comment thread means choosing which comments to include [14, 73], how to order those comments [40], and when to present them [49]. Such curation is useful when the chronological order of a comment is less relevant than its content (e.g., demoting profanity [14, 40] or highlighting political opinions [52, 53]).

Much of the work on comment curation has focused on how a reader's agreement or disagreement with the content presented affects their willingness to read or engage with it [55, 59, 74]. This is a tough sell, as most people prefer agreeable content most of the time [42, 53]. When people encounter content that challenges their own position, people may downvote it [10, 40], request fact-checks [36], or actively avoid it [20, 23]. This tendency to avoid disagreeable content raises questions about whether, when people do participate in discussions that feature positions contrary to their own, they are more likely to go off topic or introduce new topics, i.e., be less coherent.

In this paper, we extend this line of work to explore how curating comments around particular positions on a policy issue might affect new posters' willingness not just to reply but to reply coherently, increasing the chance that their comment furthers the discussion. We present results from an experiment that asked Amazon Mechanical Turk (AMT) crowd workers ("Turkers") to consider a proposed policy amendment to the AMT participation agreement to offer *partial payment* for rejected work. This proposal was presented in the context of an online discussion where comments were curated to prioritize arguments *for* or *against* the policy. We collected participants' initial positions of *support* or *opposition* to the policy as well as their comments and examined how agreement with the

position in the curated comments correlated with their own comments' coherence with existing topics.

Overall, participants were less likely to add comments that cohere with existing discussion topics when the thread curation disagreed with their initial perspective. However, this was largely driven by people who disagreed with the proposed partial payment policy, who were especially unlikely to contribute comments that cohere with existing topics when seeing a thread that prioritized support for partial payment. By contrast, people who agreed with the proposed partial payment policy were more likely to add topic-coherent comments regardless of whether the curated comments were for or against partial payment. This *asymmetric* relationship between comment curation and coherence with opposition in a policy discussion suggests that designers of both discussions and discussion forums need to consider factors beyond whether a person agrees with a particular position when considering how to support effective participation.

COHERENCE AND CURATION IN POLICY DISCUSSIONS

In a discussion, topics advance as participants reply and respond to each other along a common thread of subjects [4, 27, 67]. In this context of analysis, coherence is a function of how recent comments remain on the same topics introduced by the existing comments, which “seed” the discussion [67]. Without a coherent discussion of the pros and cons of a policy topic, it is impossible for a deliberating group to carefully weigh a policy issue [6, 26, 66].

Policy deliberation scholars have developed a few research methods for studying how groups of people talk with each other during a discussion about policy or civic issues [4]. For example, the Discourse Quality Index (DQI) is a communication coding scheme that is used to understand a policy discussion in terms of the *speeches* that people make during a discussion and how others might respond (e.g., with interruptions, counter-argument, or incivility) [65]. While the DQI is useful for studying the range of positions and level of respect during a policy discussion, a single speech may incorporate multiple topics to present a cohesive argument.

As an alternative, policy deliberation scholar Stromer-Galley [66, pg. 9] has developed a communication coding scheme to study policy discourse at the *thought* level of analysis: *A thought is defined as an utterance (from a single sentence to multiple sentences) that expresses an idea on a topic. A change in topic signaled a change in thought.* Stromer-Galley and Martinson [67] expand on the definition of topic, to characterize thoughts that add new topics to the discussion, versus thoughts that address the materials that establish the policy issue discussion (called “structuring topics”) or thoughts that address topics that emerge through the ongoing exchange (called “interactional topics”). Stromer-Galley and Martinson [67, pg. 201–205] apply what they refer to as a *dynamic topic analysis* to measure *coherence* with the interactional topics, by tracking whether new thoughts add to or divert from topics already seeded in the discussion.

We chose these methods of characterizing and measuring coherence for multiple reasons. First, Stromer-Galley [66] has

been applied in research revealing a lack of coherence in online discussion forums [28, 31]. Second, we found that the analytic granularity of distinguishing between coherence with the structural versus interactional topics was useful in our experimental design, which controls both the structuring and available interaction topics in the seeded discussion thread. Third, when people do post their thoughts to an online discussion forum, they are often in the form of comment(s), which felt closer to Stromer-Galley's definition of a thought than the DQI's notion of delivering a speech.

Coherence is one of a much broader set of concerns around deliberative discussion [4]. In this paper, we zoom in on coherence with the interactional topics because *talking-with*, and not *-past*, others in discussion is a fundamental precursor to deliberation [6, 26, 33] that is rare in online discussions [12, 64, 72], even more so when people disagree [42].

Curation and Disagreement

Disagreement is useful for small group and public decision-making. To quote John Stuart Mill's argument for why groups should not ignore opposition, “*If the opinion is right, they are deprived of the opportunity of exchanging error for truth; if wrong, they lose what is almost as great a benefit, the clearer perception and livelier impression of truth produced by its collision with error*” [50]. However, the individual experience of disagreement in a group can lead to feelings of threat [22, 42, 63], and people's reactions to these feelings can negatively affect the group [51, 41, 69]. This tension between value to group and threat to individual arises in a number of computer supported cooperative work (CSCW) contexts where people conflict with each other [19].

One potential lever designers have to manage disagreement and encourage engagement is comment curation: they can choose which comments are displayed, when, and in what order. Common strategies include showing the most recent comments, the most popular comments, and recommending personalized comments people would prefer to read [1, 53]. Many of these strategies, notably the popularity-based and personalized algorithms, tend to give people more of what they already like [57].

In this article we examine how coherence in an online policy discussion is affected by curating the discussion to promote either pro or con statements about a policy [40, 53]. Thread curation is particularly important when there are many [40] (and redundant [35]) comments in the discussion. In a policy context, thread curation can also be applied as a civic engagement lever to expose people to different views of an issue [42, 37, 52, 70, 64].

However, there is a tension in just how much opposition to present [53] and how its presentation affects a person's willingness to express their view [55, 59, 74]. This is especially risky for curation strategies that favor one side of a position over another, as might happen when trying to choose comments based on agreement or disagreement with a given participant's position, or to ensure that a particular view is heard. Further, even position-neutral strategies such as chronological order might naturally lead to situations where the discussion appears

to be tilted toward one side or another, simply because the most recent subset of comments tend to agree [40].

Curation and Cognitive Dissonance

Managing the amount of disagreement present in a curated comment thread is important because it is easy for people to avoid disagreement online [42]. While online discussion can provide people with an opportunity to form community around shared values [15, 30], properties of digital environments also enable people to stay silent among the “invisible audience” of a policy discussion [5, 25, 60].

There are various reasons why people remain silent in the face of opposition. When a person’s views are challenged they can experience *cognitive dissonance*, which can be unsettling and elicit an avoidance response [20, 23, 54], as people generally prefer not to be challenged [53]. Many people also feel unable to argue their positions, either due to a lack of training in argumentation, lack of leisure to study a particular policy matter [33, 63], or social risks of stating a position publicly [56, 64, 69].

These factors can discourage engagement with discussion topics when people see opposition in the thread, but might also encourage alternate forms of engagement that actively avoid or reduce the dissonance, such as by up/down voting comments [10, 40] or issuing fact-check requests [36]. Coe et al. discuss how people use such lightweight discussion system features in place of explicit disagreement: “[...] users often used this thumbs up/down metric in place of expressing explicit agreement/disagreement within the text of a comment.” [10, pg. 676]

This leaves open the question of when people do add a comment to a discussion thread that prioritizes views counter to their own, whether their tendency toward avoidance translates to responses that are less coherent with the existing discussion. Drawing on cognitive dissonance theory, we argue that participating in a comment thread prioritizing “agreeable” content [53] will be more pleasant [20, 23] than one that presents disagreement, and that people will be more willing and able to coherently engage an existing discussion that is on comfortable, familiar ground. Further, we would expect people who disagree to tend to change the topic in order to reduce conflict between the expressed positions and their own.

Hypothesis: *Contributions to a policy discussion are more likely to cohere when participants are exposed to a thread that prioritizes comments that match their initial position.*

METHOD

We examine the relationship between comment curation, level of agreement, and coherence in an online policy discussion via an experiment with Turkers participating in a policy discussion about the AMT participation agreement.

Interface Design

We used a discussion forum interface modeled after RegulationRoom, a platform for civic engagement in public policy-making [61]. The interface included two panels: a summary of the AMT Participation Agreement and proposals to amend

it on the left, and the comment discussion thread with a comment box on the right (Figure 1). Like RegulationRoom, the interface did not include *up/down* voting or other lightweight mechanisms for engaging with the content as our focus was specifically on topical coherence of comments rather than other behaviors.

To set basic interface design elements we first prototyped the interface. We varied the placement of the comment text box (above or below the discussion thread) and the length of the discussion thread (short, with 3 seeded comments, or long, with 20 seeded comments that required scrolling) and tested these design variations in a pilot HIT.

A total of 408 Turkers accepted the pilot HIT, of whom 292 completed it (72%). Participants averaged 35 years old and about half identified as female; about 80% were U.S.-based. Participants were randomly assigned to one of the four conditions varying the comment box position and the discussion thread length. Participants were more likely to enter a comment when exposed to the longer discussion thread (OR 9.914, $p < 0.001$). We found no effect of the comment box position on the likelihood to enter a comment. We decided to place the comment text box below the thread (as in Figure 1) based on eye-tracking research about how people read and skim articles online [18] with the hope that it would increase the likelihood of reading and engaging with other comments.

Materials

We developed the policy information presented in the discussion interface based on a summary of the AMT Participation Agreement around rejected “Human Intelligence Tasks” (HITs) posted by Requesters. A key concern for Turkers is whether a Requester accepts their work on a HIT [47], as the AMT Participation Agreement grants this power to Requesters with no recourse for workers.

We chose this specific policy topic because it has a direct impact on Turkers’ everyday lives [46, 32, 34, 47] and therefore increases the ecological validity of the study. Based on suggestions by Turkers in a prior study [47], we proposed two changes to the policy. In the first, *partial payment*, Turkers would be paid for parts of the work that were considered acceptable by the Requester. In the second, *second chance*, Turkers would have the opportunity to fix their errors in a rejected HIT.


We constructed the experiment so that the policy summary material (left pane of Figure 1) presented both partial payment and second chance; however, the comments that were seeded into the discussion thread (right pane) were exclusively about partial payment. We did this because there was more disagreement about the partial payment proposal, and because by focusing the discussion topics on partial payment, we were able to easily identify when new topics or structuring topics, like second chance, were introduced to the discussion.

To populate the discussion thread, we selected 20 comments contributed in the prior study [49], half in favor of and half opposed to partial payment. We then chose three comments pro- and three anti-partial payment as the focus comments that would be initially visible, according to the experimental

Pre Survey / Information / Discussion / Post Survey / Exit Next Stage

The Mechanical Turk Participation Agreement

An Online Discussion about Crowd Labor Market Policy



The Amazon Mechanical Turk (AMT) Participation Agreement defines the relationship between Amazon, Turkers and Requesters. The following summarizes several sections of the AMT Participation Agreement that relate to what happens when a HIT is rejected. As a reference to the policy, the summary includes links that will point you to relevant sections of the *AMT Participation Agreement* (2 Dec 2014)—for example, the following is a link to the section *Requesters in General* [§3.a].

1 | When HITs are rejected

The AMT Participation Agreement sets up two options for what happens when a HIT is completed by a Turker: (1) the Requester approves the work, and AMT automatically processes a payment to the Turker; or (2) the Requester rejects the work, and the Turker is not paid and their Approval Rating is damaged [§3.a, §3.b]. Regardless of whether the HIT is accepted or rejected, when a HIT is posted to AMT, Amazon charges the Requester a fee for their HIT request [§4.a]. However, Amazon explicitly takes a hands-off approach when there are disputes [§2, §3.f].

There are no standards for Accepting or Rejecting a HIT. The only statement about acceptance/rejection standards in the Agreement is [§3.a]: "If a Requester is not reasonably satisfied with the Services, the Requester may reject the Services." Without a definition of what is acceptable, some Requesters have used their power to define what is reasonable to [steal work](#). As Turker work is "work made for hire", Turkers give up their rights to the work as soon as it is submitted [§3.b]. Turkers use browser-based tools, like Turkopticon, to avoid risky HITs, but Turkers are limited in how they can protect themselves during and after a HIT [§3.b(i-iv)].

Turkers have proposed that better acceptance/rejection standards would improve the AMT labor market.

Comments (20)

PLEASE USE APPROVE OR REJECT TO GET PAID. IF YOU REQUEST THE WORK AND ARE NOT AWARE OF IF WHAT IS ON THE LIST IS MET BY BOTH PARTIES BUT THE REQUESTER IS UNSATISFIED THE TURKER GETS PAID 50% AND HIS/HER GENERAL RATING IS NOT DAMAGED.

@whiteReindeer Sat 8:47am

The Turker should own the rejected work. If the Requester wants ownership of the work, they should not reject it. If it's poor work that can be salvaged, the Turker most likely was not required for service in the first place and should not have been asked to waste their time on that HIT. Since they used their time to complete the HIT when they could have been doing other work that would have earned them wages, they should be paid or the work should be lost by the Requester.

@redCorn Sat 8:50am

Yes there should be clearer standers, but i agree that partial payments sounds pretty good. Maybe half?

@indigoRat Sat 9:05am

Leave a comment

Comment

Figure 1. The experiment interface, showing the policy summary on the left and the discussion thread on the right. The three comments that are initially visible and closest to the textbox are controlled to be either pro- or anti- the partial payment proposal.

condition. We added timestamps to make the discussion look recent and assigned pseudonyms (a concatenated color and animal, e.g., @blueMonkey) to each seed comment.

Participants and Recruitment

The HIT description recruited Turkers to test the user interface of a new online discussion forum platform. To attract viewpoints from a broad audience of Turkers, we did not restrict access to the HIT (e.g., to Turkers from specific countries or with specific levels of experience).

A total of 201 Turkers accepted the HIT, with 147 completing it (73%). On average, participants were 36 years old, about half identified as female, and 77% were U.S.-based. Turkers were paid \$3 for their participation; the average time to completion was 17 minutes, resulting in a pay rate of about \$10 per hour, a bit above the local state minimum wage. This payment structure adheres to the *WeAreDynamo* guidelines for Fair Payment in Academic Research¹.

Procedure

Upon accepting the HIT, participants were presented with a pre-survey that asked about their Turkling experience, variables we used as controls in our quantitative analyses. They were also directed to select a pseudonym similar to those in the seed comments, using a random name generator that concatenated colors with animal names, e.g., @blueMonkey.

Prior to entering the discussion interface, participants were informed that the discussion would be “about what happens when a HIT is rejected” and that as part of the experience they will “have an opportunity to take part in the discussion.” We informed participants that the intent is to help resolve a lack

¹http://wiki.wearedynamo.org/index.php/Guidelines_for_Academic_Requesters

of consensus among Turkers around the proposals that was observed in prior research [47]. They were asked to rate their initial position toward the two policy proposals on separate 5-item scales, from strongly disagree to strongly agree.

Participants were then placed in the discussion forum. They were able (but not required) to read a summary presentation of the relevant part of the AMT Participation Agreement and a description of the policy options, read a set of comments seeded in the simulated discussion, and add comments of their own. Participants were required to spend a minimum of one minute in the experiment interface; the average dwell time was 4.9 minutes (SD 3.6).

Curation Conditions

Each participant was randomly assigned to one of the following two conditions, each presenting the same twenty seeded comments, but sorted so that the first three comments emphasized different views toward *partial payment* (PP).

- *Pro-PP*: Three seed comments ordered closest to the comment text box presented support for partial payment.
- *Anti-PP*: Three seed comments ordered closest to the comment text box presented opposition to partial payment.

The specific comments for each condition are presented in Table 1. We randomized the order of the three comments to control for order effects and separately randomized the order of the other seventeen comments. We realize that binary categorizations as pro and anti (or agreement and disagreement) are simplifications, and that real policy discussion and positions are often more complicated. However, most prior research and many real discussion contexts do have this binary flavor, so we adopt it as well; we will return to it in the discussion.

The curated comments for each condition were selected because they share not only a similar position (Pro- or Anti-PP), but to the extent possible, similar topics in the discussion. The Pro-PP comments relate to the Partial Payment Amount (Topic 4 in our analysis; see *Coding for Topic Coherence*). The Anti-PP comments relate to the Hands-Off Labor Market (Topic 5).

While the comments are different in their position on partial payment and topic, we did not control for other characteristics of the comments (e.g., character length, expressiveness). Without an *a priori* argument about how positions on questions about the AMT participation agreement would affect Turker responses, we chose to expose Turker participants to these positions in the unaltered words of other Turkers.

Ethical considerations

Tasking AMT Turkers to discuss topics related to the AMT Participation Agreement is a familiar research context for studying systems that support policy engagement and deliberation [37, 38, 49, 62]. While this context is convenient, it requires Turkers to respond to their unequal economic position in the AMT labor market [32, 34, 47]. Unlike traditional labor markets, it is not clear how to address crowd work labor disputes through existing regulatory authorities [22], and unlike other social platforms (e.g., Reddit [8]), Turkers are not well positioned to effect change in AMT [62]. Performing research in this experimental context therefore has special ethical circumstances that need to be considered.

We received IRB-approved informed consent from all participants and compensated their time based on Turker approved standards for academic research [62]. We also implemented several Turker-supported best practices for HIT design [48], as it is important to remember that Turkers participate as part of a *task* they perform for a *reward*. In addition to taking these measures to treat participants *fairly*, we also worked with the community manager at *TurkerNation* [46] to develop the specific policy language for the study to make the content of the policy proposals relevant and engaging to the participating Turkers. Finally, we indicated that our research group is not associated with Amazon and that the purpose of the experiment was purely for research.

DATA ANALYSIS

Coding for Topic Coherence

The twenty seed comments were chosen to cover six topics identified based on 1092 Turker comments from a prior study [47] using an affinity diagramming analysis process:

1. *HIT Design*: Unclear instructions or acceptance standards and technical errors should result in partial payment
2. *Requester Communication*: Lack of Requester-to-Turker communication
3. *Turker Quality*: Low quality Turker work should not be paid (e.g., completed too quickly, robot accounts)
4. *Partial Payment Amount*: Proposes an amount or scheme for implementing partial payment (e.g., 10%, 25%, 50%)

5. *Hands-Off Labor Market*: Amazon's "hands-off" approach to the labor market (e.g., partial payment could lead to more rejections or low quality work)
6. *HIT Specific Policies*: Different protocols for different tasks (e.g., Turkers should own or receive a base payment for rejected creative work)

To identify when a comment made by a participant cohered with topics in the discussion, we used a coding scheme based on Stromer-Galley's definition of topic coherence [66, 67]. Two coders independently categorized each comment as either "new topic" or assigned a set of Topic ID numbers (1-6) identifying the topics referenced by a comment. The two coders trained initially with a set of 95 comments, resolving disagreements during the training period. Training continued until the Cohen's Kappa score for inter-rater reliability was above 0.8 and then the coding was tested on a holdout set of 95 comments. The final Cohen's Kappa score was 0.85.

The following is a sample participant comment from the current study that coheres with Topic 4 (Partial Payment Amount): "*I think partial payment should be more like 85% rather than 10%. If you only get 10% for partial payment, then I'd probably rather just redo the HIT.*" [P69]²

As an example of a comment that *did not* cohere with the seeded comments: "*I liked the idea of a second chance better than partial payment. I would like the chance to fix my mistake (if I make one). I'm honest. When I answer surveys, I read every question. I don't randomly just choose answers.*" [P14] This comment does not address the seeded topics, as it raises the *second chance* proposal which was excluded from the discussion thread. The response is somewhat related to Topic 3 (Turker Quality), though it does not speak to the specific concern that offering partial payment encourages low quality work. Unlike the prior example, the response does not provide any contextual markers that connect it to any existing interaction topic, such as the brief comparison of 10% vs. 85% that indicates that the prior comment coheres with Topic 4.

Metrics

Response Variable

The response variable used the above coding scheme to examine if a comment coheres with topics in the discussion or not.

- *Coherence (relating to existing topics)*: A hand-coded binary variable at the comment level capturing whether the comment coheres to one of the six topics addressed by the existing seeded comments.

Independent Variables

Independent variables were based on the experimental conditions and initial survey responses to the partial payment proposal.

- *Curation Condition (Pro-PP, Anti-PP)*: Captures whether participants were exposed to a discussion prioritizing comments that were pro- or anti-partial payment.

²Comments are associated with a unique identifier of the participant ranging from P1 to P147.

Pro-Partial Payment Condition	Anti-Partial Payment Condition
#995: “I mean, for work of creative nature, a base pay should be fixed. If the requester keeps and uses the work, he should pay more.”	#976: “No. Turker won’t get any partial payment. If he completes the hit with prescribed instructions ,then he will get full pay otherwise rejection.”
#1094: “perhaps there should be a template list of general criteria that every requester and turker must be aware of. If what is on the list is met by both parties but the requester is unsatisfied the turker gets paid 50% and his/her general rating is not damaged.”	#1119: “Allowing partial payment is a slippery slope, since some requesters would simply reject and give partial pay to almost everyone, citing the quality of their responses or whatever. What we need is real moderation from Amazon when there’s real abuse of the system, instead of telling us it’s between us and the requester and not their problem.”
#1136: “I believe that Turkers should receive atleast 25% of the task (if less than \$5.00) or 10% (if more than \$5.00) if it is rejected. However, they would need to have atleast shown effort and not just sped through the task. I’ve spent quite some time on a few tasks only to be rejected for something that was not clearly stated in the rules or was completely false. I believe their should atleast be an appeal system.”	#1342: “There should absolutely be clearer standards for rejecting hits and those standards should be put forth to the worker up front. Workers should be able to discuss why the hit was rejected and also able to make a case for any problem or mistake made. Unless the requester can prove that a worker was clearly just hurrying through I think a rejected work should be paid in full. If we start accepting partial payments for rejected work it will lead to requesters looking for anything to reject and then paying less than they had advertised. It could be a sticky downward spiral.”

Table 1. Comments selected to emphasize alternate views toward *partial payment* (PP) in the thread curation experimental conditions.

- *Initial Position (Support, Neutral, Oppose)*: Participants who rated their position toward partial payment as strongly agree or agree were coded as *support*; those who rated as strongly disagree or disagree were coded as *oppose*; others were coded as *neutral*.

For modeling “simple agreement” (see Table 4) we combine the Curation Condition with Initial Position into a single Matching Preference variable.

- *Matching Preference (True, False)*: Captures whether the participant’s Initial Position matched the Curation Condition (i.e., Support x Pro-PP or Opposed x Anti-PP). For this analysis of simple agreement, we removed participants with a “Neutral” view.

Control Variables

At the participant level, we controlled for participants’ self-efficacy, geographic location, and their past experience with rejections.

- *Self-efficacy*: Eight scale items of generalized self-efficacy and confidence in one’s own abilities and skills [9] were averaged into a single measure (Cronbach’s $\alpha = 0.93$). In prior research, newcomers to an online policy discussion with high assessments of their own self-efficacy contributed comments that were longer and more responsive to the policy topics [49].³
- *Country*: A binary variable coded as 1 for United States-based participants and 0 for others.

³We chose to use generalized self-efficacy as opposed to a context specific self-efficacy measure because research shows feelings of ability can translate across contexts [2]. Further, our task required multiple specific efficacy constructs (e.g., reading efficacy, writing efficacy, political efficacy); if we were to choose one, it is unclear which would be the most appropriate to measure, and measuring several would introduce extra burden on participants.

- *Rejected HITs*: In the pre-discussion survey, participants estimated the total number of HITs they have had rejected. We centered and standardized this variable, such that a one unit increase in Rejected HITs reflects a one standard deviation increase in the variable.

At the comment level, we controlled for comment length.

- *Total Words*: Total number of words in a comment, centered and standardized in the same way as Rejected HITs.

Statistical Models

As some participants made multiple comments, we treated *participant* as a mixed-effects nesting variable to account for non-independence. Mixed-effects logistic regressions were used to predict topic coherence at the comment level, as a binomial distribution was appropriate for the binary response variable. Model-level significance was evaluated using the log-likelihood ratio test, the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC).

The model coefficients are interpreted as the expected change that each independent variable contributes to the logits of the response variable. In the findings we exponentiate the logits to present the odds ratios. Odds ratios can be interpreted as the change in the response variable expected from a one-unit increase to an independent variable, holding others constant.

However, when evaluating the effect of an interaction, the coefficient estimated for the interaction is added to the main effect of the interacted variable. The combined effect of the interaction can then be exponentiated to present the effect as an odds ratio, i.e., $\exp(\text{main effect} + \text{interacted effect})$. After a model is fit to the data, the model can be used to estimate the expected likelihood of the dependent variable at various levels and combinations of the coefficients—these expected values are in terms of predicted marginal means. We use a Tukey-based pairwise comparison of the expected marginal

Participation	Count
Accepted HITs	201
Completed HITs	147
Commented	139
Total Comments	155
Control Variables	
Self-Efficacy (Mean, SD)	2.7 (0.46)
Country: US-based (Count, Percent)	122 (82.9%)
Rejected HITs (Mean, SD)	84.54 (258.43)
Total Words (Mean, SD)	296.8 (239.41)
Initial Positions for Partial Payment	
Support	53 (33.1%)
Neutral	39 (28.1%)
Oppose	55 (38.6%)

Table 2. Descriptive statistics capturing participation in the experiment, control characteristics, and details about participant initial position for partial payment prior to the discussion stage of the experiment.

	Pro-PP	Anti-PP	Total
Participants	74	73	147
Total Comments	81	74	155
Coherent	38	34	72

Table 3. Descriptive statistics capturing the count of participants, comments, and coherent comments in each curation condition (Pro-PP, Anti-PP). Almost 95% of participants contributed a comment.

means to examine the interactions within the models (using a 95% confidence interval).

FINDINGS

Descriptive Overview

Table 2 presents the descriptive statistics for the study. As part of our question is whether initial position might influence behavior, we first confirm that the distribution of initial positions was not significantly different between the curation conditions: $\chi^2(2, 147) = 3.013, p = 0.2217$.

Table 3 reports the number of participants, comments, and coherent comments by condition. Although the task instructions explicitly did not require participants to leave a comment, most participants did so, with just under half cohering with existing topics (46.4%). We found no significant difference in the likelihood to make a comment by condition: $\chi^2(1, 147) = 1.9e-29, p = 1$. Therefore, we focus on the likelihood to cohere with the existing topics within the discussion thread.

Agreeing with Curated Position Increased Coherence

The data support our hypothesis that presenting participants with content that matches their initial position increases the likelihood of contributions that cohere with the existing discussion topics. Comments made by participants whose initial position matched the curation condition were 2.970 times more likely to cohere with the existing discussion ($p < 0.05$, see Table 4). This mirrors HCI research about recommending content that is agreeable [53] and similar to what a user already likes [1, 57].

	Coherence	
	Est (SE)	OR
Curation Condition and Initial Positions		
(Intercept)	-2.691 (1.54)	0.067 .
Matching Preference	1.088 (0.44)	2.970 *
Control characteristics		
Self-Efficacy	0.691 (0.54)	1.995
Country: International	-0.338 (0.57)	0.713
Rejected HITs	-0.328 (0.71)	0.720
Total Words	0.408 (0.18)	1.503 *
Log likelihood	-70.02 (df=7)	

Table 4. Fixed-effects logistic regression predicting the likelihood that participants will engage specific seeded discussion topics when their expressed preference matches the curation condition (i.e., Pro-PP x Support, Anti-PP x Opposed).

p-value significance: *** 0.001; ** 0.01; * 0.05; . 0.1;

Opposed Positions were Overall Less Coherent

However, the significant *Intercept* in Table 4 and relatively weak significance of the *Matching Preference* covariate indicate that the model is missing a good deal of variance. Thus, we next examine a model that distinguishes between participants' initial positions (Table 5). This analysis shows that beyond agreement or disagreement, initial position matters. Participants *Opposed* to partial payment were significantly less likely to post responses that cohere with the discussion topics than those who *Support* it (OR = 0.173, $p < 0.01$). Further, *Opposed* participants in the Anti-PP condition, who see comments they agree with, were significantly more likely to cohere than in Pro-PP: $\exp(-1.753 + 2.224) = 1.600$ OR ($p < 0.01$). This finding aligns with the primary argument, that agreement and coherence go hand in hand, although initial position also helps predict coherence.

Effects of Initial Position were Asymmetrical

Because interaction effects where there are multiple levels can be tricky to evaluate, we next applied a Tukey-based pairwise comparison of each of the variable levels (e.g., comparing coherence likelihood between initial Neutral and Support positions, while keeping the curation condition fixed at Anti-PP). Figure 2 graphically depicts the analysis as a predicted marginal means interaction plot between discussion curation (Pro-PP and Anti-PP) and initial position for partial payment (Support, Neutral, Oppose).

Two main points emerge out of this analysis. First, the Pro-PP curation condition generated significantly more *coherent* contributions from those with *Neutral* or *Support* initial positions compared to those *Opposed*, 7.054 times ($p < 0.01$) and 5.292 times, ($p < 0.05$) respectively, while the Anti-PP curation condition did not show this difference. Second, the main driver of this effect is those with an *Opposed* initial position, as their contributions were significantly less likely to cohere, by 0.228 times ($p < 0.01$) in the Pro-PP condition versus Anti-PP. This difference between the conditions does not occur in other cases.

DISCUSSION, IMPLICATIONS, AND LIMITATIONS

At a high level, the findings confirm our hypotheses that curating a discussion thread to match a participant's preferences

<i>Coherence</i>		
	<i>Est (SE)</i>	<i>OR</i>
<i>Curation Condition</i>		
(Intercept)	-1.215 (1.26)	0.296
Anti-PP	-0.850 (0.60)	0.427
<i>Initial Position for Partial Payment</i>		
Neutral	0.208 (0.65)	1.232
Opposed	-1.753 (0.64)	0.173 **
<i>Control characteristics</i>		
Self-Efficacy	0.615 (0.42)	1.850
Country: International	-0.597 (0.53)	0.550
Rejected HITs	-0.568 (0.76)	0.566
Total Words	0.443 (0.17)	1.557 **
<i>Interaction (Curation Condition x Initial Position)</i>		
Anti-PP x Neutral	0.043 (0.92)	1.286
Anti-PP x Opposed	2.224 (0.85)	1.600 **
Log likelihood	-94.41 (df=10)	

Table 5. Fixed-effects logistic regression predicting the likelihood that participant responses will cohere with specific seeded discussion topics. Initial Position terms are by comparison to Support for partial payment and Curation features are by comparison to the Pro-PP condition, both of which are baselines in the Intercept. The odds ratios reflect the exponentiation of the estimates for each feature. p-value significance: * 0.001; ** 0.01; * 0.05; . 0.1;**

increases the likelihood of contributions that cohere with the existing discussion. This observation extends existing research about how people prefer content that agrees with their preferences [42, 53, 57], by demonstrating how curation can affect whether new contributions cohere with or diverge from the existing discussion. However, the effect of comment curation and participant preference was asymmetrical: people who support partial payment were much more likely to engage coherently with the conversation when they saw comments curated to present the Anti-PP position, compared to people opposed to partial payment who saw the Pro-PP curation condition.

This observation implies that curating primarily for position consistency led us (and presumably, will lead others) to downplay other aspects of comments and contributors that might affect people’s coherent engagement with diverse perspectives on a policy issue. For instance, Munson and Resnick found that reaction to individual items in a curated list of news articles depended not only on that single item but also on the others surrounding it [53]. Moving towards implementation, the discussion platform *ConsiderIt* [37] allows an individual user to adopt a variety of arguments both for and against any given policy decision.

This brings back to our choice of binarizing positions and comments on the policy into a pro versus anti framing, as such work shows how, in practice, reactions rarely occur in response to a single, isolated policy proposal. (Indeed, for that reason we originally planned to examine both second chance and partial payment rather than focusing on just one proposal. However, that plan was ultimately discarded, both due to the complexity of the analysis and the limited comparability to prior work.)

Taken together, these results and work suggest that research and design around engaging with disagreement would benefit

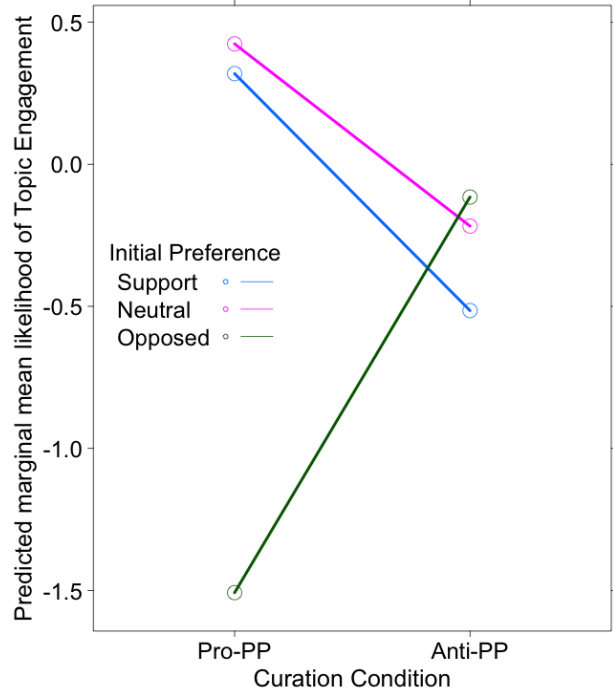


Figure 2. Predicted marginal means interaction plot between discussion curation (Pro-PP and Anti-PP) and initial position for partial payment (Support, Neutral, Oppose). Linear predictions are given on a log scale and reflect the estimated probability of a topic engagement.

from more nuanced views and analyses of how and why people agree or disagree with the positions proposed. Tools such as ConsiderIt, or argument analysis techniques from the area of natural language processing, could lead in fruitful directions for better supporting contentious policy discussions.

Curating for openness rather than agreement?

Our results also raise an important caveat around comment curation: curating comments risks systematically excluding viewpoints. The general versions of this are familiar: popularity tends to curate for and perpetuate majority opinion; personalization tends to curate for agreement, which may increase engagement but primarily with like-minded people [57]. In our case, although people in all conditions were equally likely to contribute a comment, people who were Opposed to partial payment were much less likely to cohere with the current discussion (especially when presented with Pro-PP arguments)—and contributions seen as off-topic are often ignored.

In this experiment, one possible driver of the asymmetry was a difference in the specificity of the featured Pro-PP and Anti-PP comments. Pro-PP was more specific around implementation decisions—i.e., setting the right level of partial payment—while Anti-PP comments were more general, describing how the partial payment proposal would have damaging effects on workers by giving HIT Requesters more room to reject requests.

We posit a parallel to how high fidelity interface prototypes tend to elicit comments about specific design elements, versus napkin sketches that give more room for considering the overall interaction [7]. Proposals that focus attention on specific implementation details will tend to concentrate attention on those details, arguably leaving less room for coherent discussion on topics where there is disagreement than more open discussions of the policy context. Opponents seeing Pro-PP curation may also have felt like they were joining the planning committee for a distasteful proposal—an unlikely scenario for effective, coherent contributions—while supporters seeing Anti-PP curation may have felt like their opinion on the issue still contributed to its deliberation.

In an attempt to integrate topic coherence [67] with common definitions of deliberation, such as careful weighing of diverse perspectives [6, 26, 66], we suggest that discussion and deliberation moderators might want to encourage (support) coherent disagreement: *a thread of comments that consistently contribute to a careful weighing of differing perspectives on a topic*. Designing to support coherent disagreement might imply highlighting content that leaves more room for debate (just as good interviewers, and bad lawyers, ask open-ended questions). How well this strategy for promoting coherent disagreement would work is an open question—choosing a controlled, one-shot experiment on one topic means we cannot make strong claims to generality or ecological validity—as is the question of what properties of a comment would invite openness.

Still, the idea has potential and is worth further study, both at the level of individual comments and of groups of them. Online discussion moderation means managing a stream of comments, often one at a time [14, 39, 40], or engaging specific comments to learn more about the experience of specific commenters [61, 21]. Supporting coherent disagreement might mean curating (or moderating) sets of comments based on characteristics of the group: expressing a range of positions, possessing topical coherence as a group, affording overall openness to discussion, representing a diversity of stakeholders, and so on.

Changing curation strategies and metrics over time

Another consideration is that the goals of a deliberation change over time—and another reading of the curated comments in this study is that the Pro-PP condition presented the deliberation as farther along than the Anti-PP condition. As the state of a deliberation transitions from investigating a common problem, to eliciting a range of potential ideas, to critiquing the ideas and refining them into a single proposal, each transition starts to impose constraints on the discussion topics. Thus, the discussion naturally tends to narrow through proposal development, making it harder to make coherent contributions, especially for those who disagree with the fundamental approach rather than with some implementation detail.

The shifting nature of policy discussion and deliberation group tasks over time suggest that curation strategies and metrics should likely change with them. We focused on coherence in this paper because it is understudied [24] and because *talking-with* the existing discussion is a common value in many policy

discussions [12]. However, our measure of coherence is only appropriate for some goals. In our definition, the opposite of being coherent was not “incoherent”, but could include introducing novel topics and ideas; such divergent thinking has real value at many stages of many group processes [17, 16].

More generally, work around curation—including this paper—has tended to focus on the specific problem of curation to support engagement with disagreement. This is an important problem to be sure, but is a small part of a much wider range of deliberation desiderata [6, 26, 33]. One might curate (or moderate) for many of these values, including civility and quality [68], soliciting both objective and subjective descriptions [45], supporting both social and task processes [26], eliciting both logical arguments and situated experiences and stories [58, 61], and so on. Curating for disagreement might come at the expense of other goals such as civility, social affect, or solidarity; putting disagreement at the center leaves these other important concerns at the margin.

CONCLUSION

Online discussions about policy are shaped by the ways that contributors dynamically add to, expand on, or divert attention away from existing topics in the discussion. Here we operationalize this dynamic by measuring coherence with the interactional topics in an online discussion, and predict higher levels of coherence when participants are exposed to a comment thread that prioritizes positions of the policy that match their own preference. However, we observed an asymmetric relationship between preference and curation, which implies that curating primarily for or against consistency with a poster’s current position may overlook other important aspects of comments, including their openness to coherent deliberation, that might affect people’s willingness to engage coherently with diverse perspectives on a policy issue.

ACKNOWLEDGEMENT

We thank Daniel Wu for developing the infrastructure for this research and Sara Chau for her support through the analysis. Thanks also to Kristy Milland and the Cornell eRulemaking Initiative (CeRI) team for their support developing the policy context for the AMT Participation Agreement discussion. This work is supported by NSF HCC 1314778.

REFERENCES

1. Lars Backstrom, Jon Kleinberg, Lillian Lee, and Cristian Danescu-Niculescu-Mizil. 2013. Characterizing and curating conversation threads: Expansion, focus, volume, re-entry. In *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, New York, NY, 13–22. DOI: <http://dx.doi.org/10.1145/2433396.2433401>
2. Albert Bandura. 1997. *Self-efficacy: The exercise of control*. W. H. Freeman and Company, New York, NY.
3. Laura W. Black. 2012. Blog, chat, edit, text, or tweet? Using online tools to advance adult civic engagement. *New Directions for Adult and Continuing Education* 2012, 135 (2012), 71–79.

4. Laura W Black, Stephanie Burkhalter, John Gastil, and Jennifer Stromer-Galley. 2010. Methods for analyzing and measuring group deliberation. *Sourcebook of political communication research: Methods, measures, and analytical techniques* (2010), 323–345.
5. danah boyd. 2007. Why youth (heart) social network sites: The role of networked publics in teenage social life. *MacArthur foundation series on digital learning—Youth, identity, and digital media volume* (2007), 119–142.
6. Stephanie Burkhalter, John Gastil, and Todd Kelshaw. 2002. A conceptual definition and theoretical model of public deliberation in small face-to-face groups. *Communication Theory* 12, 4 (2002), 398–422.
7. Bill Buxton. 2010. *Sketching user experiences: getting the design right and the right design*. Morgan Kaufmann.
8. Alissa Centivany and Bobby Glushko. 2016. Popcorn aates good: Participatory policymaking and Reddit’s amageddon. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI16)*, San Jose, CA.
9. Gilad Chen, Stanley M. Gully, and Dov Eden. 2001. Validation of a new general self-efficacy scale. *Organizational research* (2001). <http://journals.sagepub.com/doi/abs/10.1177/109442810141004>
10. Kevin Coe, Kate Kenski, and Stephen A. Rains. 2014. Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication* 64, 4 (8 2014), 658–679. DOI: <http://dx.doi.org/10.1111/jcom.12104>
11. Todd Davies and Reid Chandler. 2011. Online deilberation design: Choices, criteria, and evidence. *Democracy in motion: Evaluating the practice and impact of deliberative civic engagement* (2011), 103–131.
12. Richard Davis. 1999. *The web of politics: The Internet’s impact on the American political system*. Oxford University Press.
13. Fiorella De Cindio and Cristian Peraboni. 2010. Design issues for building deliberative digital habitats. *Online Deliberation* (2010), 41.
14. Nicholas Diakopoulos and Mor Naaman. 2011. Topicality, time, and sentiment in online news comments. In *CHI’11 Extended Abstracts on Human Factors in Computing Systems*. 1405–1410.
15. Carl DiSalvo. 2009. Design and the construction of publics. *Design issues* 25, 1 (2009), 48–63.
16. Steven Dow, Julie Fortuna, Dan Schwartz, and Beth Altringer. 2011. Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2807–2816. <http://dl.acm.org/citation.cfm?id=1979359>
17. Steven Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel Schwartz, and Scott Klemmer. 2010. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction (TOCHI)* 17, 4 (2010). <http://dl.acm.org/citation.cfm?id=1879836>
18. Geoffrey B. Duggan and Stephen J. Payne. 2011. Skim reading by satisficing: Evidence from eye tracking. *Proceedings of the SIGCHI Conference on* (2011). <http://dl.acm.org/citation.cfm?id=1979114>
19. Steve M. Easterbrook, Eevi E. Beck, James S. Goodlet, Lydia Plowman, Mike Sharples, and Charles C. Wood. 1993. A survey of empirical studies of conflict. In *CSCW: Cooperation or Conflict?* Springer, 1–68.
20. Andrew J. Elliot and Patricia G. Devine. 1994. On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of personality and social psychology* 67, 3 (1994), 382. <http://psycnet.apa.org/journals/psp/67/3/382/>
21. Dmitry Epstein and Gilly Leshed. 2016. The magic sauce: Practices of facilitation in online policy deliberation. *Journal of Public Deliberation* 12, 1 (2016). <http://www.publicdeliberation.net/jpdhttp://www.publicdeliberation.net/jpd/vol12/iss1/art4>
22. Alek Felstiner. 2011. Working the crowd: Employment and labor law in the crowdsourcing industry. *Berkeley Journal of Employment and Labor Law* (2011), 143–203.
23. Leon Festinger. 1962. *A theory of cognitive dissonance*. Vol. 2. Stanford University Press.
24. Dennis Friess and Christiane Eilders. 2015. A Systematic review of online deliberation research. *Policy & Internet* 7, 3 (9 2015), 319–339. DOI: <http://dx.doi.org/10.1002/poi3.95>
25. Susan R. Fussell and Leslie D. Setlock. 2014. Computer-mediated communication. *The Oxford Handbook of Language and Social Psychology* (2014), 471.
26. John Gastil and Laura Black. 2008. Public deliberation as the organizing principle of political communication research. *Journal of Public Deliberation* 4, 1 (2008).
27. Erving Goffman. 1976. Replies and responses. *Language in society* 5, 03 (1976), 257–313.
28. Todd Graham. 2012. Beyond “political” communicative spaces: Talking politics on the Wife Swap discussion forum. *Journal of Information Technology & Politics* 9, 1 (1 2012), 31–45. DOI: <http://dx.doi.org/10.1080/19331681.2012.635961>
29. Ali Gürkan, Luca Iandoli, Mark Klein, and Giuseppe Zollo. 2010. Mediating debate through on-line large-scale argumentation: Evidence from the field. *Information Sciences* 180, 19 (10 2010), 3686–3702. DOI: <http://dx.doi.org/10.1016/j.ins.2010.06.011>
30. Maarten Hajer. 2003. Policy without polity? Policy analysis and the institutional void. *Policy sciences* 36, 2 (2003), 175–195.

31. Daniel Halpern and Jennifer Gibbs. 2013. Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior* 29, 3 (2013), 1159–1168. DOI: <http://dx.doi.org/10.1016/j.chb.2012.10.008>
32. Lilly C. Irani and M. Silberman. 2013. Turkopticon: Interrupting worker invisibility in Amazon Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 611–620. <http://dl.acm.org/citation.cfm?id=2470742>
33. Christopher F. Karpowitz and Chad Raphael. 2014. *Deliberation, democracy, and civic forums: Improving equality and publicity*. Cambridge University Press.
34. Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 1301–1318.
35. Mark Klein. 2011. How to harvest collective wisdom on complex problems: An introduction to the MIT Deliberatorium. *Center for Collective Intelligence working paper* (2011).
36. Travis Kriplean, Caitlin Bonnar, Alan Borning, Bo Kinney, and Brian Gill. 2014. Integrating on-demand fact-checking with public dialogue. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14*. ACM Press, New York, New York, USA, 1188–1199. DOI: <http://dx.doi.org/10.1145/2531602.2531677>
37. Travis Kriplean, Jonathan Morgan, Deen Freelon, Alan Borning, and Lance Bennett. 2012a. Supporting reflective public thought with ConsiderIt. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. 265–274.
38. Travis Kriplean, Michael Toomim, Jonathan Morgan, Alan Borning, and Andrew Ko. 2012b. Is this what you meant?: Promoting listening on the web with Reflect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1559–1568.
39. Cliff Lampe and Paul Resnick. 2004. Slash (dot) and burn: Distributed moderation in a large online conversation space. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 543–550.
40. Cliff Lampe, Paul Zube, Jusil Lee, Chul Hyun Park, and Erik Johnston. 2014. Crowdsourcing civility: A natural experiment examining the effects of distributed moderation in online forums. *Government Information Quarterly* 31, 2 (4 2014), 317–326. DOI: <http://dx.doi.org/10.1016/j.giq.2013.11.005>
41. Edward J Lawler. 2001. An affect theory of social exchange1. *Amer. J. Sociology* 107, 2 (2001), 321–352.
42. Azi Lev-On and Bernard Manin. 2009. Happy accidents: Deliberation and on-line exposure to opposing views. *Online deliberation: Design, research and practice* (2009), 105–122.
43. Matthew T. Loveland and Delia Popescu. 2011. Democracy on the web. *Information, Communication & Society* 14, 5 (8 2011), 684–703. DOI: <http://dx.doi.org/10.1080/1369118X.2010.521844>
44. Rousiley C. M. Maia and Thaianie A. S. Rezende. 2016. Respect and disrespect in deliberation across the networked media environment: Examining multiple paths of political talk. *Journal of Computer-Mediated Communication* 21, 2 (3 2016), 121–139. DOI: <http://dx.doi.org/10.1111/jcc4.12155>
45. Meethu Malu, Nikunj Jethi, and Dan Cosley. 2012. Encouraging personal storytelling by example. *Proceedings of the 2012 iConference* (2012). <http://dl.acm.org/citation.cfm?id=2132309>
46. David Martin, Benjamin V. Hanrahan, Jacki O’Neill, and Neha Gupta. 2014. Being a Turker. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 224–235.
47. Brian McInnis, Dan Cosley, Chaebong Nam, and Gilly Leshed. 2016a. Taking a HIT: Designing around rejection, mistrust, risk, and workers’ experiences in Amazon Mechanical Turk. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2271–2282.
48. Brian McInnis and Gilly Leshed. 2016. Running user studies with crowd workers. *Interactions* XXIII, 5 (2016), 50. <http://dl.acm.org/citation.cfm?id=2968077>
49. Brian McInnis, Elizabeth Murnane, Dmitry Epstein, Dan Cosley, and Gilly Leshed. 2016b. One and done: Factors affecting one-time contributors to ad-hoc online communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 609–623.
50. John Stuart Mill. 1975. *On liberty* (1859). Norton.
51. Richard L. Moreland and John M. Levine. 1982. Socialization in small groups: Temporal changes in individual-group relations. *Advances in experimental social psychology* 15 (1982), 137–192.
52. Sean A. Munson, Stephanie Y. Lee, and Paul Resnick. 2013. Encouraging reading of diverse political Viewpoints with a browser widget. In *ICWSM*. <http://dub.uw.edu/djangosite/media/papers/balancer-icwsm-v4.pdf>
53. Sean A. Munson and Paul Resnick. 2010. Presenting diverse political opinions: How and how much. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1457–1466.
54. Diana C. Mutz. 2008. Is deliberative democracy a falsifiable theory? *Annual Review of Political Science* 11, 1 (6 2008), 521–538. DOI: <http://dx.doi.org/10.1146/annurev.polisci.11.081306.070308>

55. Elmie Nekmat and William J. Gonzenbach. 2013. Multiple opinion climates in online forums: Role of website source reference and within-forum opinion congruency. *Journalism & Mass Communication Quarterly* 90, 4 (2013), 736–756. <http://journals.sagepub.com/doi/abs/10.1177/1077699013503162>
56. Elisabeth Noelle-Neumann. 1974. The spiral of silence: A theory of public opinion. *Journal of Communication* 24, 2 (1974), 43–51. <http://onlinelibrary.wiley.com/doi/10.1111/j.1460-2466.1974.tb00367.x/full>
57. Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin, UK.
58. Francesca Polletta and John Lee. 2006. Is telling stories good for democracy? Rhetoric in public deliberation after 9/11. *American Sociological Review* 71, 5 (2006), 699–721.
59. Pablo Porten-Cheé and Christiane Eilders. 2015. Spiral of silence online: How online communication affects opinion climate perception and opinion expression regarding the climate change debate. *Studies in Communication Sciences* 15, 1 (2015), 143–150. <http://www.sciencedirect.com/science/article/pii/S1424489615000211>
60. Jenny Preece, Blair Nonnecke, and Dorine Andrews. 2004. The top five reasons for lurking: Improving community experiences for everyone. *Computers in human behavior* 20, 2 (2004), 201–223. <http://www.sciencedirect.com/science/article/pii/S0747563203000876>
61. Cynthia R. Farina, Dmitry Epstein, Josiah B. Heidt, and Mary J. Newhart. 2013. Regulation Room: Getting “more, better” civic participation in complex government policymaking. *Transforming Government: People, Process and Policy* 7, 4 (2013), 501–516.
62. Niloufar Salehi, Lilly C. Irani, Michael S. Bernstein, Ali Alkhatib, Eva Ogbe, Kristy Milland, and Clickhappier. 2015. We Are Dynamo. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, New York, New York, USA, 1621–1630. DOI: <http://dx.doi.org/10.1145/2702123.2702508>
63. Lynn M. Sanders. 1997. Against deliberation. *Political theory* (1997).
64. Bryan C. Semaan, Scott P. Robertson, Sara Douglas, and Misa Maruyama. 2014. Social media supporting political deliberation across multiple public spheres: Towards depolarization. *Proceedings of the 17th* (2014), 1409–1421. <http://dl.acm.org/citation.cfm?id=2531605>
65. Marco R. Steenbergen, André Bächtiger, Markus Spörndli, and Jürg Steiner. 2003. Measuring political deliberation: A discourse quality index. *Comparative European Politics* 1, 1 (3 2003), 21–48. DOI: <http://dx.doi.org/10.1057/palgrave.cep.6110002>
66. Jennifer Stromer-Galley. 2007. Measuring deliberation’s content: A coding scheme. *Journal of public deliberation* 3, 1 (2007).
67. Jennifer Stromer-Galley and Anna M. Martinson. 2009. Coherence in political computer-mediated communication: Analyzing topic relevance and drift in chat. *Discourse & Communication* 3, 2 (5 2009), 195–216. DOI: <http://dx.doi.org/10.1177/1750481309102452>
68. Abhay Sukumaran, Stephanie Vezich, Melanie McHugh, and Clifford Nass. 2011. Normative influences on thoughtful online participation. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, New York, New York, USA, 3401. DOI: <http://dx.doi.org/10.1145/1978942.1979450>
69. Cass R Sunstein. 2006. Deliberating groups versus prediction markets (or Hayek’s challenge to Habermas). *Episteme* 3, 03 (2006), 192–213.
70. W. Ben Towne and James D. Herbsleb. 2012. Design considerations for online deliberation systems. *Journal of Information Technology & Politics* 9, 1 (2012), 97–115.
71. Jing Wang, Clement T. Yu, Philip S. Yu, Bing Liu, and Weiyi Meng. 2012. Diversionary comments under political blog posts. In *Proceedings of the 21st ACM international conference on Information and knowledge management - CIKM '12*. ACM Press, New York, New York, USA, 1789. DOI: <http://dx.doi.org/10.1145/2396761.2398518>
72. Scott Wright, Todd Graham, and Daniel Jackson. 2016. Third space, social media and everyday political talk. *The Routledge companion to social media and politics*. New York: Taylor & Francis (2016), 74–88.
73. Scott Wright and John Street. 2007. Democracy, deliberation and design: The case of online discussion forums. *New media & society* 9, 5 (2007), 849–869.
74. Thomas Zerback and Nayla Fawzi. 2016. Can online exemplars trigger a spiral of silence? Examining the effects of exemplar opinions on perceptions of public opinion and speaking out. *New Media & Society* (2016). <http://nms.sagepub.com/content/early/2016/01/22/1461444815625942.abstract>
75. Amy X. Zhang, Lea Verou, and David Karger. 2017. Wikum: Bridging discussion forums and wikis using recursive summarization. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 2082–2096.