

Is This What You Meant? Promoting Listening on the Web with Reflect

Travis Kriplean¹, Michael Toomim¹, Jonathan Morgan², Alan Borning¹, Amy J. Ko³

Computer Science & Engr.¹, Human Centered Design & Engr.², Information School³
University of Washington, Seattle

ABSTRACT

A lack of support for active listening undermines discussion and deliberation on the web. We contribute a design frame identifying potential improvements to web discussion where listening more explicitly encouraged in interfaces. We explore these concepts through a novel interface, Reflect, that creates a space next to every comment where others can summarize the points they hear the commenter making. Deployments on Slashdot, Wikimedia's Strategic Planning Initiative, and a local civic effort suggest that interfaces for listening may have traction for general use on the web.

Author Keywords

Web, Discussion, Listening, Deliberation, Grounding

ACM Classification Keywords

H.5.0 Information Interfaces and Presentation: General

INTRODUCTION

To accomplish collective goals, people must discuss what to do, how to do it, and who will do it, whether it is an open source project, a faculty meeting, or a government planning process. We are seeing a shift toward participatory engagement, in which organizations are exploring new ways of including their constituencies in decision making and people increasingly expect such opportunities [4]. Discussions can be merely consultative, as with Facebook's gathering of feedback on terms of service changes and U.S. President Obama's gathering of citizens' ideas for open governance. Or it can be at the heart of organizational culture, such as in Wikipedia where participants strive for community consensus in all decisions [5, 20]. We believe these deliberative efforts to tap the creative energy, emotion, and intelligence of many are needed to improve our collective ability to confront public problems.

Unfortunately, public discourse on the web is often perceived as inflammatory and hyperbolic [15], which constrains the potential of web-based public discussion for informing decisions. We argue that constructive web discussion is elusive partially as a result of the lack of attention to supporting listening. Interfaces designed for listening might nudge people

to establish more common understanding, even when the discussion is heated. When participants establish mutual understanding, they are more likely to recognize the legitimacy of that person's perspective [27], and the quality of their opinions are often improved [9]. Our goal in this paper is to draw attention to the generative potential of a deeper understanding of listening on the web, and illustrate it through the design, deployment, and evaluation of a novel interface.

We contribute a design-oriented theoretical frame for supporting listening in web interfaces. Second, we contribute a novel interface, Reflect, which embodies the listening design frame. Reflect introduces a second dimension into online comment boards, adding a backchannel for people to demonstrate evidence of listening by restating the points that they hear the commenter making (Fig. 1). Inspired by Wikipedia, Reflect nudges people to strive toward a neutral reflection of what others are saying, even if there is disagreement. Our third contribution is a presentation of Reflect's deployment in three diverse contexts: a high-volume technology news discussion site, summarization by facilitators during Wikimedia's Strategic Planning Initiative and during a local civic engagement effort. Our results suggest that restatement is an activity in which discussion participants will voluntarily engage.

INTERFACES FOR GROUNDING AND LISTENING

Active listening is integral to communication. Listeners must provide evidence to speakers to show that they are being attentive and understand [11]. Listeners operate a backchannel where they nod, say "uh huh", tilt their heads, finish sentences, and preface responses with a simple restatement of what the speaker said [17]. These explicit acts of listening helps speakers "debug" their messages, as well as provide evidence that they are being recognized and heard [33]. It also helps listeners demonstrate good faith as conversation partners; partners who provide more feedback through the backchannel are perceived as more patient, polite, and attentive [36]. This process of trading off speaking and listening to establish mutual understanding is called *grounding* [11].

In contrast, web interfaces have typically not explicitly supported backchannels for others to demonstrate evidence of understanding. Consider two ends of the spectrum of current (implicit) support for listening. First, threaded forums allow replies, which can be appropriated by users to listen, such as by rephrasing what is being said or asking for clarification. But mixing replies and acts of listening can quickly make a discussion difficult to follow if many people are involved. Moreover, there is usually an expectation that comments add something new to the conversation. Second, many discussion systems allow participants to vote on their favorite utterances,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI'12, May 5–10, 2012, Austin, Texas, USA.

Copyright 2012 ACM 978-1-4503-1015-4/12/05...\$10.00.

Con:: The constitutional right to bail is critical for many poor defendants.

 James B But the only people that this amendment would apply to--those charged with crimes that might put them away for life, AND who have demonstrated a propensity for violence--are already having their bail set at extremely high levels. Which means that only the poor defendants are waiting around in jail, the rich ones get bailed out.

 Vin Hill This amendment would allow wealthy defendants to be denied bail. The applicable laws

Figure 1. The first use of Reflect in the wild, in the Living Voters Guide [23]. In this exchange, Vin restates a point he thought James made, and then replies. James later returned and clarified his point to correct Vin, illustrating an ordering challenge with asynchronous listening.

such as with “like” or “recommend” buttons. These buttons overload two functionalities: they perform an important listening functionality, in that they signal to the commenter that they have been heard; but they also serve to *judge* the statement. Thus, these mechanisms are not pure listening mechanisms, but also contain an implicit “I agree” reply. This does not serve those who disagree with what is being said but yet wish to recognize and appreciate the speaker.

Overall, web interfaces implicitly privilege speaking over listening, creating a feedback chasm that may hinder formation of common ground. Prioritizing speaking over listening may bias toward emotionally-charged but shallow interactions, in turn affecting who is willing to participate. We argue that interfaces can be designed that help nudge people toward more reflective interactions by emphasizing the common experience of listening. We hypothesize the following four potential outcomes of supporting listening:

Listening interfaces enable people to see evidence that they are being heard, improving their communication satisfaction and willingness to participate. Consider reactions to the frustration of not being heard. One option is forcefully restating a point, further deteriorating the situation. A second reaction is to disengage. Being responded to is one determinant of whether people continue contributing in an online forum [2], and the fundamental dynamic may be being heard.

Listening interfaces empower participants to use and hone their active listening skills to guide the discussion and demonstrate their own worth. Acts of listening impact the direction that a conversation takes [11, 33], and help to demonstrate to others that the listener is a valuable participant. Consider the motivations for active listening: first, a listener might try to demonstrate understanding before responding (such as by restating), a commonly recommended technique for effectively interacting with someone about a controversial or complex issue (e.g., marriage counseling or dispute mediation in Wikipedia [6]). Second, listeners might provide feedback that helps teach speakers how to better frame their points. Third, listening can draw other’s attention to a statement’s importance (e.g., if the point reflects their own personal beliefs [15]). Fourth, someone might articulate understanding in order to better interpret the intended meaning.

By showing active listeners’ demonstrations of understanding (or misunderstanding), listening interfaces help other discussants make sense of what is being said and

- Are you saying that you believe this law would not be applied to wealthy and poor defendants **equally?** *Vin Hill*

 - I meant that the CURRENT law does not treat the wealthy and the poor equally, since the wealthy can afford levels of bail that the poor cannot. *James B.*

- [What do you hear James saying?](#)

- [What do you hear Vin saying?](#)

why. In an examination of how speakers and listeners work together to coordinate a conversation, Kraut [22] found that while an active listener is the one whose understanding of what a speaker is saying is most positively affected when interacting with the speaker, eavesdroppers also benefit. This effect is familiar: consider a large meeting where a colleague succinctly rephrases your long and convoluted expression of an important point. Your colleague has demonstrated that she has understood, and also helped everyone else understand better. These are acts of facilitation that lead others to say, “Ahhh, I see!” Listening interfaces may help elicit, capture, and expose these “ahh” moments for future participants.

Listening interfaces help establish an empathetic normative environment. Even simple changes to a commenting interface can impact the thoughtfulness of comments, leading to the establishment of commenting norms as participants calibrate to the observed behaviors of prior participants [31]. Recent research in neuroscience indicates that the likelihood of empathetic engagement, e.g. via perspective taking, is sensitive to experimental manipulation [24], suggesting that we might successfully embed cues for listening in our interfaces. If the interface can encourage some users to listen, others may follow, helping to establish constructive communicative norms. Through their visible structuring, web interfaces can continually remind people to actively listen, even when the discussion is heated, and re-present these acts to other participants to bolster the listener’s conversational status.

We have not tested all of these hypotheses. Rather, we contribute the speaking/listening theoretical frame as a useful metaphor for considering new design directions and guiding future empirical investigations, and demonstrate its productivity with the interface presented next. In reading this paper, it is important to keep in mind that the speaking/listening frame is a metaphor. There are significant differences between the asynchronous character of the web discussions we target and the synchronous dyadic face-to-face communication upon which much of the research cited thus far is based. For example, we know that the introduction of asynchrony, even short delays, can disrupt speaker/listener coordination [34]. Some may argue that reading/writing is a more accurate metaphor for asynchronous web discussions. However, the reading/writing metaphor only indirectly captures the communicative functionality that listening highlights: timely, fine-grained feedback for communicators. Reading implies interaction between a reader and text, not in-

teraction with the writer, whereas writing encompasses both raising new points and addressing points raised by the other.

REFLECT: LISTENING THROUGH RESTATEMENT

The design space for listening is wide. For example, a minimal listening mechanism might simply be an “I hear you” button, which does not support grounding but might lead to some of the interpersonal benefits discussed. A non-textual listening mechanism for supporting grounding might have users annotate text with an emoticon that conveys their sentiment when reading that piece of text (smiling, grimace, quizzical). Emoticons from multiple people could be aggregated and visualized on mousehover, engaging themes raised in the prior section. A maximal listening interface would have users restate the points they hear the commenter saying, motivated by a useful practice when in an emotional argument.

In this paper, we experiment with this latter strategy of restatement. We present a novel interface, Reflect, that makes a lightweight change to online comment boards: to the right of every comment, visitors are invited to restate points they hear the commenters making (Fig. 2). These restatements, each an act of listening, are publicly displayed in a bulleted list. Reflect provides an opportunity for the original commenter to respond to each bullet point to verify its accuracy and clarify if necessary. Reflect thus establishes a listener’s backchannel that engages the theoretical propositions laid out earlier. First, commenters can learn how others are interpreting their statements and clarify misunderstandings, supporting one iteration of grounding. Second, listeners can gain status as productive listeners. Third, other participants can scan the bullets to better understand what is being said and identify important points, assuming that listeners are more likely to restate points that they find important (a form of “read wear” [21]). And finally, in contrast to most comment boards, a Reflect-enabled site strongly suggests to a visitor that listening is a primary mode of participation, perhaps helping to establish a normative environment through which common ground is more likely to be found. By carving out a prominent space for restatements in a comment board, we gain leverage for inquiring about what would happen if we make active listening a first-class activity in web discussions.

Scenario of use and design rationale

Imagine that you have just come to a discussion about whether the City of Seattle should eliminate an ordinance requiring real estate developers to provide for future tenant parking. You start browsing the comments, glancing over the first one and then read the summary points others have written. When you hover over a summary bullet, the relevant text in the comment that the bullet refers to is highlighted, allowing you to jump between the commenter’s own words and the listener’s interpretation (Fig. 2.1). This particular summary is interesting because it helps you better understand the point the commenter was making about the free-market implications of your idea. You mark the restatement as “Helps shed light on what the commenter was trying to say” (Fig. 2.2).

You recognize the next commenter, Lisa, who added the helpful restatement. You note that no one has restated any of her points, so you read her comment carefully. You add a bullet point so that she knows someone is listening and so that her

comment is not lost in the mix. After clicking “Add a point that Lisa made”, you type in your interpretation of her main argument in the 140-character limited text field (Fig. 2.3). It causes you to stop and think for a while, which is often not how you usually skim comments. After hitting submit, you are asked to connect the point that you summarized to the part of Lisa’s comment to which it refers (Fig. 2.4). After clicking two sentences, which are subsequently highlighted, you click done. The bullet shows up next to the comment.

You scroll down to a comment you posted earlier. Several bullet points have been added. You recall receiving notification emails, but had not had time to check it out then. The interface presents you with the option of verifying the accuracy of each bullet point, or if it is even a restatement at all. You can also add a short clarification if needed. The first bullet is accurate. You click “yes” and a checkmark appears next to the summary. The second bullet is inaccurate. You click “no” and clarify what you were trying to say (Fig. 2.5). The clarification shows up underneath the bullet point (Fig. 2.6).

Articulating the work of comment summarization. There are other ways in which the task of restating or summarizing the comment could be broken up besides short bullet points. For example, our first design used a wiki next to the comment. However, in early testing, this proved problematic. People felt that with a wiki they must commit to summarizing the whole comment (many of which can be quite long). We moved to bullet lists and discovered some additional affordances: (1) including more people in the listening and grounding activity in a natural way, (2) enhanced support for clarifying points, (3) facilitating multiple interpretations of the same text, (4) easier for commenters to respond to the summary and clarify their points, (5) easier to skim the summaries and connect them to the comments, (6) the ability to identify similar points being made in other comments, and (7) the ability for listeners to highlight the points they found most salient in what the commenter said.

Connecting summary text to comment text. After someone summarizes a point that a commenter makes, we ask him or her to click on the sentences in the original comment where the point was being made. There were two reasons for doing this. First, accountability: a listener explicitly commits to exactly where a commenter made the point they believe them to be making. For those who are maliciously misinterpreting the message or deliberately antagonizing the commenter, it adds a step whereby they are forced to essentially acknowledge their maliciousness. Second, readability: this data is used later to aid readers in figuring out the relevant parts of the comment to which the summary refers. If there is a breakdown in grounding between speaker and listener, the speaker may be able to more easily see where the misinterpretation occurred. Others can judge whether the listener accurately represented the point. If someone maliciously writes a bullet, it’s relatively easy to verify what the sentences actually said.

Speakers are privileged to verify summaries. Because Reflect places the bullet points next to a comment, an inaccurate bullet might overly influence what others believe the commenter was saying. Therefore, Reflect enables speakers to verify a summary bullet point and clarify when needed, which is prominently displayed to other readers underneath the bul-

Readers

1 Hovering over bullets highlights relevant text in the comment.

2 Readers can positively or negatively evaluate bullet points.

Listeners

3 Any reader can become a listener by restating a point they hear the commenter making in 140 characters or less.

4 After restating, a listener connects their bullet point to the relevant sentences in the comment.

Speakers

5 A commenter can verify the accuracy of a restatement of their comment, and clarify if necessary.

6 Verifications are prominently displayed to other readers.

Lisa Russell

This idea has already been put into action by the Seattle City Council with disastrous results. If you want to eliminate parking requirements, then make each person who moves into a building with no requirements sign an affidavit that they will not park their car on the street- if they won't sign then they can't live there. The current result of this idea is that developers make more money because they haven't provided parking- and neighborhoods lose because new people come in and park on the street.

1 Readers hear Lisa saying...

- Instead, the city should implement better neighborhood coverage of public transit and zipcar. *ballardborn*
- Seattle City Council already tried this before, and it didn't end well *Katie*
- What do you hear Lisa saying?

2 Does Katie's summary... elegantly distill meaning? uncover a good point? clarify the message? provoke unnecessarily? umm, its not a summary

3 What do you hear Lisa saying?

4 The lack of parking motivates outsiders to use scarce parking spaces *Henry*

Done cancel 103 summaries, not replies

← Please click the relevant sentences

Done cancel

5 The lack of parking motivates outsiders to use scarce parking spaces *Henry*

Is this accurate?

- Yes
- Maybe, but I should clarify
- No, its not a summary

6 The lack of parking motivates outsiders to use scarce parking spaces *Henry*

clarification: yes, but its not just outsiders, its new residents whose developers did not provide them parking

Figure 2. Mechanics of the Reflect interface.

let point. However, Reflect does not entrust speakers with the ability to directly edit or delete the bullet points. Otherwise they may censor other people's valid interpretations. Privileged responses address the tension between allowing the speaker to repair breakdowns with giving listeners the power to express their interpretations. These verifications thus facilitate public grounding of expressed meaning.

Community evaluation. Not every restatement is created equal. Some restatements might highlight a point deeply buried in a long comment, or distill the essence of a long comment. Or bullets might be antagonistic or used to reply. Reflect therefore enables other participants to recognize good restatements and flag poor ones. Aside from potentially removing malicious restatements, this can help to demonstrate for participants the purpose that good restatements might play.

Implementation. Reflect is an open source jQuery Javascript library. It requires a configuration object specifying CSS selectors for key DOM elements. On page load, the core library can then wrap the Reflect markup around comments. A server-side web service for fetching and storing Reflect data must also be implemented. The service can also carry out other tasks like email notifications. Reflect implementations are publicly available for Wordpress, Drupal, and Mediawiki.

Related design work

Reflect explicitly supports the process of grounding described by Clark's contribution theory [11]. HCI researchers have

drawn upon this theory as an analytical lens to understand the affordances of the communication medium (e.g. [19]), how communication strategies change under pressures of scale [35], or informing the design of systems supporting distributed teamwork [14]. To our knowledge, Reflect is the first interface that makes the process of content grounding a first-class activity in web-based discussions. Other interfaces afford restatements, such as annotation and anchoring tools [7, 10], but do not call out this affordance explicitly. The only work we have found that explicitly supports restatement is the classroom tool ThoughtSwap [16]. Students submit ideas in response to a prompt, and then other students can pull ideas out of the "hat" and re-present those ideas. Ultimately, Reflect is meant to trigger deeper reflection about what others are saying. In this sense, one of the most related systems is MetaViz [3]. MetaViz attempts to trigger critical thinking by using computational metaphor identification to expose the political metaphors that participants are drawing upon in a blog. ConsiderIt [23] is also highly related. ConsiderIt enables people to create pro/con lists for a given issue, but also include points others have written into their own list. This inclusion functionality is similar in spirit to a Reflect summary.

There is a rich line of research that helps users make sense of conversations. First, a number of interfaces help users navigate threaded discussions, such as visualizing social and semantic associations [28, 30], or focus+context interfaces for reading threads [32]. Others have attempted to overcome

sensemaking challenges by moving away from threaded comment boards entirely. A recent exemplary system is OpinionSpace [18]. OpinionSpace plots on a two-dimensional map the individual comments in a web forum, based on the commenters' responses to a short value-based questionnaire. By navigating this space, readers are better able to seek out a diversity of comments as well as prime themselves for engaging the perspective of someone with different values. Argumentation visualization systems, a particularly well explored class of systems, also provide an alternate structure to threaded forums. These systems require users to break down their opinions according to argumentation schemas drawn from linguistic theory or elsewhere, and incorporate them into a graphical visualization of the issues, arguments, positions, and so forth (see [8] for an overview). Argumentation visualization has proven to have serious limitations in practice, including the difficulty of learning the formalized schemas, breaking up narratives into fine chunks, and agreeing on classifications [29], often requiring trained facilitators for effective use [13]. Overall, these interfaces can help set the context for listening, but they do not go beyond the typical rating of comments when it comes to engaging with what is being said.

Finally, researchers have created tools for extracting take-aways from discussions, such as by creating discussion-level summaries. Ackerman and colleagues have tackled this goal with expert systems for post-processing discussions [1, 26]. Underlying this work is the stance that the ideal case would be automatic summarization (see e.g. [26], p. 139). While automatic summarization would be very useful, this ideal precludes the potential interpersonal benefits of listening.

FIELD DEPLOYMENTS

To provide insight into how Reflect might be appropriated in real discussions, we report on three deployments: (1) augmenting four stories posted to the popular technology news discussion site Slashdot; (2) use by the Wikimedia Foundation to summarize input they received during a worldwide Strategic Planning Initiative; and (3) in a local civic engagement initiative (CCF), providing a means for the leader to show participants on their blog that he was listening to them.

We divide our deployments into two contexts of use: *community news discussion* (Slashdot) and *community deliberation* (Wikimedia, CCF). Because of the size of Slashdot's readership, the Slashdot deployment helps us most in establishing whether and how Reflect will likely be used by a wide public. The other deployments are more limited because details of each deployment limited the size of the user base. The value of these other deployments is to illustrate additional possible use cases for community deliberation.

Community News Discussion

Two fundamental questions are whether online discussants will (1) actually use the Reflect functionality and (2) whether they will use it to listen to each other, rather than troll or reply to commenters. These questions are essential in determining if Reflect, and listening interfaces more broadly, might have traction for broad public use. If it does, then pursuing this line of inquiry further may be fruitful, both in terms of creating new designs for listening as well as rigorously examining whether the hypothesized outcomes laid out earlier hold.

Table 1. Absolute counts of four contribution types in the Slashdot data (and number of unique users contributing each).

	Comments	Bullets	Ratings	Replies
Story 1	389 (145)	336 (113)	719 (285)	20 (13)
Story 2	112 (67)	151 (47)	335 (203)	17 (7)
Story 3	57 (37)	81 (36)	181 (129)	14 (6)
Story 4	174 (93)	166 (65)	264 (144)	12 (8)

To answer these questions, we deployed Reflect on four articles posted to Slashdot, one of the most popular news discussion sites since its founding in 1997. Slashdot's audience is primarily male professional engineers, an audience one might not expect to embrace Reflect and its emphasis on listening. If the Slashdot community uses Reflect as intended, it suggests that Reflect may be widely applicable, with the caveat that the tech savvy user base with its experience using the complex Slashdot commenting system [25] may be quicker to understand Reflect than the general public.

The opportunity to deploy Reflect on Slashdot was made possible by an introduction to Slashdot founder Rob Malda by fellow researcher Cliff Lampe.¹ From Spring to Fall 2011, we developed the Slashcode plugin and coordinated with Malda and engineer Tim Vroom. Through this process, Reflect's design evolved, including the addition of the community moderation functionality. This emerged as an important design addition because Malda, with his decade-and-a-half experience moderating online discussions, was concerned that Reflect would be used predominantly to troll or reply to commenters. We designed the bullet evaluation functionality, and then used the evaluations to deactivate bullet points that were rated negatively. Specifically, in our scheme, if more than half of the evaluations of a bullet point were negative after the first three evaluations, the bullet was deactivated and hidden. The bullet was also deactivated if the commenter labeled the bullet as "not a summary" during verification and at least one reader gave it a negative rating.

In mid-September 2011, Slashdot enabled Reflect on four stories. During the trial, **734 bullets were written by 247 discussants, an average of 1.0 bullets per comment** (Table 1). Figure 3 shows an excerpt from Story 4. The full discussion is embedded in the margin. 15.3% of commenters in the stories also created a bullet point, overall contributing 41.2% of the bullets. The remaining 58.8% of the bullets were contributed by people we may previously have labeled as "lurkers".

Did the written bullet points actually reflect the meaning conveyed in the comment? Our most important empirical results, after establishing the baseline that Reflect would in fact be used, concern the nature of the bullet points and the extent these bullets reflected back the commenter's points. Here, we did content analysis on the bullet points. This was challenging because the bullet points were deeply contextual, rich social expressions drawing on multiple levels of shared meaning: the Slashdot culture and past interactions; implicit references to the story being discussed and the preceding comments; and fairly detailed, often nuanced comments.

We coded all 734 bullet points using a standard grounded

¹All of our deployment contacts and interview participants gave us explicit permission to use their full names.

The screenshot shows a comment from user EdIII on a Slashdot story. The comment reads:

And? (Score:4, Insightful)
by EdIII (1114411) on Tuesday September 20, @10:16PM (#37463536)

Do you trust such companies with unfettered access to the entire GPS history of your vehicle?

Of course I don't. I don't own a vehicle that has the ability to be shut off remotely either, because I don't trust a company or the systems with something that important. I would not trust the electric company with my refrigerator either. The very fact the control exists with a 3rd party is unacceptable.

If you are worried about being tracked, OnStar is the least of your concerns. It applies to a single source of data that is not always with you.

Anybody that really cares about this should wonder what data is being collected with your smart phones, etc. Verizon can track you better than OnStar ever could.

Points *EdIII* makes

- You are being tracked even more closely by other companies. Why worry so much about your car doing it too? *yakatz* ✓
- If losing control of your personal data alarms you, and it should, then there are other items of concern as well. *walkerp1*
- OnStar is the tip of the iceberg. People need to be more careful with who they trust their data to and should pay attention to who has it. *Anubis IV*
- You can only trust yourself. Companies are inept, evil, or ~~too~~ profit-driven to be trusted. *mclearn*

clarification: Trust is a factor, but so is risk management. The greater the risk, the greater the level of trust that should be required. *EdIII*

• [What do you hear EdIII saying?](#)

Full discussions

Slashdot
Mediawiki
Mechanical Turk

Figure 3. Deployment screenshots, focusing on an excerpt from a Slashdot story. Zoom in to read the full discussions off to the right.

analysis through which we iteratively developed and refined a codebook and established inter-rater reliability. Ultimately, we decided to have each coder independently code every bullet and resolve discrepancies in order to yield results in which we are confident.² The codebook has two levels, shown in Table 2. The highest level is a yes/no code: Does the bullet primarily reflect back material from the comment? We achieved $\alpha = .89$ reliability for answering this question.

The codebook has five primary subcodes for bullet points that reflected meaning. First, *neutral reflections* rephrased, distilled, or elaborated upon (e.g. through examples) the respective comments' meaning. They do not go beyond the comment's meaning in any significant way. If *neutral reflection* was selected, no other code could also be selected because the purpose of this code was to capture strict restatements. Second, the *expands meaning* code was used for bullet points that read between the lines, such as by generalizing, inferring antecedents or consequences, or putting a high-level label on a phenomena the commenter described (e.g. "Social justice" when the commenter discusses income tax equalities). Third, the *meta observation* code was applied to bullet points that contextualized the meaning of the comment, such as by focusing a critical eye on the rhetorical devices employed by the comment or positions the meaning with respect to the norms of the community or other external events. Fourth, the *negative/cheeky* code was applied to bullet points that are non-neutral, usually sarcastic summaries that often mocked the ideas at play, or were overtly negative toward the comment's

content. This code helps distinguish bullets that, while reflecting meaning, do so in a potentially antagonizing manner. Fifth, the *contains replies* code was applied when a bullet, after reflecting the meaning of the comment, also inserted a response to that point.

There are three subcodes for bullets that do not reflect meaning. The *flames* code captures bullet points that simply served to personally attack the commenter via ad hominem. It does not engage the substance of the comment, but rather the commenter. The *replies* code is for bullet points that reply to the comment without reflecting back any meaning. *Offtopic* bullets are unrelated to the discussion.

To illustrate the rich range of bullet points users wrote and better demonstrate these codes, we highlight six bullet points (out of 24) created for a polarizing comment. The comment took place on a story about the U.S. federal government closing down an online poker site because it was operating as a Ponzi scheme. The comment stated: *When Mitt Romney asks, "Why punish success?", I suggest people think about [poker sites like this]. Not all businesses are scams, but the people raking in millions of dollars a year aren't earning it. They're inheriting it, winning it or stealing it, and they deserve to be taxed at a higher rate.* Consider these bullet points for this comment:

1. *Just because someone obtained their wealth legally, does not mean it is wealth they have earned or generated.* Coded as **neutral reflection**.
2. *Implication that Romney isn't so different from the people involved in the criminal scheme.* Coded as **expands** because it reads between the lines. As this example shows, bullets that expand meaning can draw an inference that does not definitively follow from the comment.
3. *Not all business is criminal, only successful business.* Coded as **cheeky, expands** because the bullet writer expands the meaning of the comment by equating "people

²The codebook was developed by categorizing 50 random bullets, then iterated on with the first, second and fourth authors as we independently coded and resolved discrepancies on additional sets of bullets. Ultimately we converged on a two-level codebook. On a random set of 100 bullets, we established high reliability on the high-level code, though reliability on some subcodes was far less reliable. Low reliability was due to difficulty identifying all the necessary cues on each code, not because we disagreed on phenomena. Thus we elected to independently code each bullet and resolve any discrepancies. Discrepancies usually resulted from one of the coders attending to a signal that had escaped the others' attention.

Table 2. Content analysis: codebook, reliability and results.

	Krippendorff's α	Slashdot	Wikimedia Strategy	Mechanical Turk
<i>Reflects back meaning</i> of the comment	0.89	69.0%	97.0%	92.0%
<i>Neutral reflection</i> without elaboration	0.73	23.7%	92.0%	65.0%
<i>Expands meaning</i> by reading between the lines	0.66	35.8%	4.0%	26.0%
<i>Meta observation</i> that contextualizes the content	0.84	12.5%	0.0%	1.0%
<i>Antagonistic or cheeky</i> reflection of meaning	0.45	11.9%	1.0%	1.0%
<i>Contains a reply</i> in addition to reflecting	0.58	3.0%	0.0%	0.0%
<i>Doesn't reflect meaning</i> contained in the comment	0.89	31.0%	3.0%	8.0%
<i>Flames</i> the commenter with an ad hominem	0.84	10.4%	0.0%	0.0%
<i>Replies</i> to the commenter without reflecting	0.68	10.2%	0.0%	8.0%
<i>Off-topic</i> and doesn't relate to the discussion	1.00	9.6%	0.0%	0.0%
<i>Repeats</i> Simply repeats part of the comment	1.00	1.9%	3.0%	0.0%

Table 3. Moderation results. The community used the bullet evaluations to sanction and hide bullets that did not reflect meaning, increasing the visual salience of listening.

	active bullets	moderated bullets
<i>Reflects back meaning</i>	442	61
<i>Doesn't reflect meaning</i>	113	118

raking in millions of dollars” with “successful business” to make it possible for the *reductio ad absurdum*.

4. *reflecting the current US topic of ‘class warfare’ related to taxing higher incomes at the same rate as lower incomes.* Coded as **meta** because it contextualizes the comment with respect to the larger social issue at play.
5. *[the commenter] is emoting as per his training and will be rewarded by his fellow jealous and equally well trained primates.* **Flames** because this is an ad hominem attack. If, however, the bullet stated ...*his training about how the rich have cheated their way to wealth and should be taxed more....*, it would have been **cheeky**, **meta** because it reflects back meaning in a negative way while also making a prediction about how Slashdotters will react.

69.0% of the bullets Reflected back meaning, indicating Reflect’s use as a listening interface. Discounting negative/cheeky bullets, 57.1% of all bullets listened in an apparently neutral fashion. Of the subcodes, “expands meaning” was the most prevalent. As coders, we found many of the bullets falling into this category to be the most useful because they often brought to light the implied meaning of an unclear, subtle, or obfuscated comment. While flaming and pure replies were common, given concerns aired by Slashdot co-founder Rob Malda, we were pleasantly surprised that they were not more frequent. The distribution of frequencies in Table 2 as well as the set of examples given above conveys a variety of rich ways in which Reflect was used to listen.

Did the community’s moderation of bullets tend to deactivate bullets that did not restate points? When enough users rated a bullet as not a restatement of one form or another, the bullet was deactivated and hidden from view for subsequent users.

24.6% of all bullets were moderated by the community and hidden from future view. Of these moderated bullets, 65.9% were not summaries (Table 3). This means that **after community moderation is factored in, 79.6% of the final set**

of publicly visible bullet points neutrally reflected meaning. Moreover, 40.4% of the moderated bullets that reflected meaning were negative/cheeky, while 51.1% of all non-reflecting bullets were deactivated. We thus conclude that the community sanctioned non-restatements, carving out a backchannel for listening.

This deployment illustrates what discretionary use of Reflect might look like in a discourse community. We found an almost one to one comment to summary ratio. Aside from raw use, we found that participants generally followed our design intentions in writing bullets that restated points, and that the community moderation further reinforced this intention. These results are particularly promising given the reputation of Slashdot as an often vitriolic community.

Community Deliberation

Communities often need to deliberate and make decisions about what to do about challenges it faces. Sometimes this takes the form of a government, non-profit, or movement organization launching a project to engage its constituents. Or it might be a bottom-up discussion started by a community member that foments a new movement or influences decision-makers to take action. These efforts at community deliberation are the types of discussions we are primarily interested in improving through listening interfaces. In this section we describe two deployments of Reflect that highlight distinct use cases for community deliberation, primarily drawing on interview data from the leaders of each initiative. Unfortunately, these deployments are not as strong as the Slashdot deployment, so the implications we can draw are more limited. Still, these results provide value because it helps designers understand more deeply, if imperfectly, some of the different contexts of use to which Reflect might be put, and highlights opportunities for future empirical work.

Summarization in Wikimedia’s Strategic Planning Initiative

Professional facilitators often need to summarize and extract themes from long discussions. This is an important but resource intensive task for effective community deliberation. Reflect may help with this process, as it allows facilitators (or participants, depending on the conditions of the deployment) to create intermediate summaries at the comment level. These intermediate summaries can then be clustered and synthesized to create a high level executive summary.

To gain insight into this use case, we were fortunate to partner with Wikimedia’s Strategy Planning Initiative (SPI).

Wikimedia, the organization behind Wikipedia and its sister projects, ran the SPI from mid-2009 to mid-2010 in an effort to gather input from volunteers worldwide about the strategic directions the organization should take for promoting Wikimedia efforts. Near the end of the initiative, Eugene Kim, the lead collaboration consultant contracted to design and lead the SPI, contacted the first author about using Reflect internally to process and summarize the input they had been receiving. During June and July 2010, five facilitators and volunteers installed a Greasemonkey script that implemented Reflect. *This meant that Reflect restatements were only visible to the people who installed the Greasemonkey script.*

These facilitators created 282 unique summary bullets to aid their summarization task. See Figure 3 to see these facilitators using Reflect. Using the same methodology developed for the Slashdot case, we coded a sample (100) of the bullet points. See Table 2. The bullets that the facilitators created were far more straightforward than in Slashdot, with 97% of the bullets reflecting meaning (and 92% pure reflections). This is unsurprising given that the facilitators decided to use Reflect with the explicit purpose of summarization. More interesting is the workflow that facilitators reported: copying and pasting bullet points into the thread level summary, and then synthesizing them. This is indicative of an important future direction for Reflect: development of and explicit support for a summarization workflow. The SPI leader's public forum post summarizes the value facilitators found in Reflect:

We used Reflect as part of Wikimedia's open strategic planning process, an unprecedented year-long effort that drew over a thousand participants. As you can imagine, continuous synthesis was a critical component of the process, and it required a Herculean effort by some of our volunteers to summarize some of our discussions. That changed when we installed Reflect. Reflect created a lightweight way for summarizing individual posts in a participatory way. Having more people summarize this way vastly increased people's abilities to synthesize long, complex threads. I believe it had an impact on the quality of discussion as well, although we didn't have a chance to thoroughly evaluate this. I think Reflect is particularly powerful exactly because it takes one – and only one – small, but very concrete step toward augmenting a conversational style that people are already familiar with. That approach is more likely to scale than approaches that attempt to take an all-encompassing approach to shaping conversations.

Listening by leaders, in Countywide Community Forum
Deliberation is often part of a decision-making process coordinated by an organization or leader. A major challenge these efforts face, particularly governments, is that constituents or community members may feel that there is no reason to participate because their voices will not be heard by those in power [12]. Reflect might help with this problem: leaders and facilitators can show that they are listening simply by restating points, without necessarily promising specific actions.

For this use case, we originally intended to deploy Reflect with the US Federal Communications Commission, which at the time was gathering public input regarding Net Neutrality legislation. The social media team at the FCC were excited

Experience writing bullet points, e.g. *Who was the primary audience for the bullet points you wrote?*

Experience having a comment they wrote summarized by someone else, e.g. *What was your reaction to having someone restate what they thought they heard you say?*

Experience reading a Reflect-enabled discussion, e.g. *Did having comments with restatements on the right change how you read comments?*

Perceptions of Reflect's affordances and limitations given professional experience, e.g. *What kinds of discussions do you think Reflect would be good for? Why? What kinds of discussions would Reflect NOT be good for? Why?*



Figure 4. Interview questions were divided into four groups. Zoom in to read the full transcripts on the right.

about using Reflect to show participants that they were listening, without needing to reply to each person. However, public outreach was canceled as the FCC director got embroiled in a political firestorm over his initial proposal. Instead, we partnered with the leader of the Countywide Community Forums (CCF), a local civic engagement effort with the mandate to stimulate discussion about issues relevant to citizens of King County. The initiative was established in 2007 with the passing of the “Easy Citizen Involvement Initiative” ballot measure. About once a quarter, CCF selects a new local issue, and provides support for citizens to host discussions at private homes, libraries and elsewhere. Surveys are distributed so that citizens can record their opinions. At the end of the round, the coordinators spin these surveys into a report which is circulated in the media and presented to elected officials.

The volunteer leader, John Spady (JS), contacted us about installing Reflect in order to support a new online outreach effort for the initiative. JS wants to try to cultivate community as much as possible around the comments the site receives, with Reflect as one part of the strategy (“*When I Reflect, I'm doing it with a sense that I want to engage the unseen readers. It's not so very much for the original author – except to try and let him/her know that someone is reading and responding to his/her comments.*”). During this short deployment, JS created 37 Reflect summary points, covering every comment, between August 2011 and September 2011. Unfortunately, an error led to data loss so the bullets could not be coded. The bullets were similar in nature to the Wikimedia deployment.

Interviews with discussion facilitators about use of Reflect

We conducted structured interviews via email with four expert users: JS from CCF, and Eugene (EK), Philippe (PB) and Kristofer (K) from the SPI. Using an email interview protocol, we asked the same four sets of questions of each participant. In the initial email, we also attached a full-page screenshot of a Reflect-enabled discussion in which they had participated so that they could better recall their experiences. These interviews are not meant to result in generalizable, objective knowledge, but rather provide insight into how experienced professionals view Reflect's affordances and limitations after using it. Fig. 4 contains the classes of interview questions, as well as an embedded image of all interview transcripts for the curious reader (with interviewee permission). We briefly draw attention to two themes from the interview results:

All participants sought to write bullets that distilled long passages into succinct statements. In response to a question about what makes a good Reflect point and how they approached writing a bullet point, JS wrote “*Short and sweet! Trying to write the least I can to capture the essence of a portion of the original document*”; PB writes “*It’s almost like writing a tweet: short, to the point, and provided the necessary info to recreate my thought process later*”; EK states “*My goal was always to highlight the essence of the post, no more, no less. I’d try to do it in a way that captured the tone of the original summary*”. K added that “*what was good with Reflect was that it was possible to strip away all the ‘proofs’ of the statement and express the idea in a single sentence, but still having the text that the bullet was derived from highlighted when a user hovered over it and thus making it possible for anyone critical of the statement to investigate the reasoning more thoroughly.*” In other words, Reflect provided an opportunity to express meaning in the absence of justification.

Reflect caused all four participants to consider more deeply the meaning the commenter intended to convey. Each participant wrote that Reflect led them to “*pause and really listen to what the other person is trying to say*” (JS). PB states that “*We tend to jump instinctively on comment threads and Reflect forced me to stop and consider the underlying meaning of what the commenter said, as opposed to just what I initially read.*” Similarly, EK writes “*Occasionally, I would read a sentence that I didn’t completely understand, and I would generally just gloss over those. You can’t do that when you’re looking to summarize.*” K adds that the highlight functionality was important to this outcome, as “*sometimes I wrote a bullet point that I couldn’t highlight, and that often meant that the bullet point wasn’t all too accurately formulated, and I had to reconsider if I actually had understood what the reader really meant yet.*” This experience of summarizing other comments was powerful enough for K and PB that they started to alter how they wrote comments so that their comments were clearer (“*I found myself editing comments afterwards for style (removing subordinate clauses, difficult punctuation, etc)*”). Overall, these responses are encouraging because they suggest that Reflect may act to counteract our tendency toward knee-jerk reactions.

FUTURE WORK AND CONCLUSION

By designing interfaces that help improve interpersonal dynamics and summarize long discussions, we hope to foster the conditions under which large community deliberations can be a more viable and effective venue through which publics can brainstorm, decide upon, and take action on shared problems. This paper has made several contributions toward this goal. It has presented a theoretically informed design frame proposing opportunities for overcoming problems with web discussion through an attention to listening. This paper also contributes Reflect, a novel interface makes the design perspective concrete. Our deployments demonstrated traction for voluntary public use and for summarizing full discussions.

Advancing the empirical agenda

There are many opportunities for advancing our empirical understanding of listening interfaces. First, longer-duration deployments can address questions about whether listening interfaces can help foster empathetic norms and characterize

whether there was a novelty effect in the nature and extent of usage we observed. Second, there are many contexts in which discussions take place. By necessity we had to be opportunistic in our deployments, as many deployment opportunities fell through. One future direction would be to deploy listening interfaces in a wide range of contexts to characterize when, how, and why they are most useful. We suspect that Reflect has the most potential for shifting patterns of dialogue in situations where (1) discussants share common social bonds such as identifying with a shared community (e.g., Slashdot), (2) discussants share a common goal (e.g., citizens communicating with government about how to address a specific issue), or (3) decision makers employ Reflect as a mechanism by which to explicitly listen to constituents. Conversely, we do not expect that Reflect alone would do much to improve discussions on news sites like the Washington Post.

Finally, early in the paper we laid out a number of potential outcomes of designing listening interfaces. There are major challenges to validating hypotheses about the effects of a discussion interface upon interpersonal experiences and the quality of discussions, since the topic and participants greatly impact the interpersonal dynamics of the discussion. In particular, the entire discussion is strongly dependent on the starting conditions, so to get more general results we need to have many different independent discussions.

To address these challenges, we have been developing a methodology using Amazon Mechanical Turk for conducting A/B tested discussion experiments without losing ecological validity. Mechanical Turk allows anyone to post small tasks to be completed for small amounts of money. “Requesters” post tasks and “workers” (or “turkers”) complete them. A community of workers has emerged, bonded by shared experiences, such as being cheated by requesters who refuse to pay for completed work or posted tasks that are part of efforts of dubious moral nature (e.g., posting fake product reviews).

The presence of a community of turkers means that researchers can start discussions about topics naturally interesting to participants, have access to a platform ideal for recruiting, and leverage an environment conducive to controlled experiments. Our method is to post tasks where workers take a survey and optionally participate in a discussion about Mechanical Turk. These workers are then randomly assigned to one of several interface conditions. Exit surveys and other techniques can be used to measure differences among the conditions. We conducted a preliminary experiment using this technique.³ There were two simultaneous, independent discussions that participants were randomly assigned to after they “accepted” the HIT. Each discussion had the same prompt: “*What should Amazon do about bad HITs [aka microtasks] and requesters?*” (Fig. 3). A self-report exit survey about how participants perceived others and the quality of the discussion more broadly captured whether there might be differences between the two conditions. We saw no significant

³We recruited over a hundred participants. 95 comments (45 replies) were added in the Reflect condition and 77 comments (28 replies) in the non-Reflect condition. In the Reflect condition, 142 summary bullet points were created (1.49 bullets per comment). Figure 3 shows a full screenshot of the Reflect condition, and Table 2 shows the result of coding a sample of 100 bullet points from the Reflect-enabled condition.

differences in the self-report results from the exit survey between conditions.

Despite attracting hundreds of participants, this only counts as a single datapoint in terms of independent discussions. But because this discussion can be seen as part of larger community dialogue, we can generate additional discussions by extracting the ideas that workers contributed and host new discussions around each one, facilitating a series of A/B-tested discussions that build toward recommendations for Amazon or even a Turker Bill of Rights. By conducting experiments on Mechanical Turk about Mechanical Turk, researchers gain an ecologically valid naturalistic deployment (participants discussed an issue they cared about, were knowledgeable about, and were not directly compensated for their participation), while still garnering some level of experimental control through the independent discussions.

Toward Building from Reflect

We argue that acts of listening can provide a foundation upon which more holistic discussion systems can be designed. Not only do acts of listening provide grist for summarization, but they also enable the creation of reputation systems that reward participants not just for speaking well, but also for listening well. To explore these concepts, we will next create a discussion system illustrating a discussion summarization workflow based on Reflect, and a reputation system that rewards participants with the ability to make longer posts when they speak well (e.g. comments others like), listen well (create positively rated Reflect bullets) and help summarize the discussion. The summarization activity will capture the most important take-aways from the discussion (e.g. actionable ideas), gradually written over time by many readers doing small tasks that are automatically combined (e.g. incorporating a set of restatements into the summary, verifying that a restated point is reflected in the summary, or improving the flow of the summary). If social infrastructure can be created for participants to engage empathetically with each other and summarize their collective thoughts, we can enable decision makers and other participants to actually engage with the points being expressed in a large discussion, opening a wide range of high-value applications.

ACKNOWLEDGEMENTS

Thanks to Eugene Kim, Rob Malda, Tim Vroom, John Spady, Deen Freelon, Cliff Lampe, Volkmar Pipek, CSCW Horizon judges, NSF grant IIS 0966929 and many others for supporting this research and advancing the ideas contained herein.

REFERENCES

1. M. S. Ackerman et al. *I-DIAG: from community discussion to knowledge distillation*, pages 307–325. Kluwer, B.V., 2003.
2. J. Arguello et al. Foundations for successful individual-group interactions in online communities. In *CHI*, 2006.
3. E. Baumer et al. Fostering critical and creative thinking about metaphor in political blogs. In *CHI*, 2010.
4. W. L. Bennett. Changing citizenship in the digital age. *MacArthur Series on Digital Media and Learning*, 2007.
5. I. Beschastnikh et al. Wikipedian self-governance in action. In *ICWSM*, 2008.
6. M. Billings and L. Watts. Understanding dispute resolution online. In *CHI*, 2010.
7. A. J. Brush et al. Anchored discussions vs. discussion boards. In *CSCL*, 2002.
8. S. Buckingham Shum. Cohere: Towards web 2.0 argumentation. In *COMMA*, 2008.
9. J. Cappella et al. Electronic dialogue during campaign 2000. *Political Communication*, 19(1), 2002.
10. E. F. Churchill et al. Anchored conversations. In *CHI*, 2000.
11. H. Clark et al. Grounding in communication. In *Perspectives on socially shared cognition*. Am. Psych. Assoc., 1991.
12. S. Coleman and J. Blumler. *The Internet and Democratic Citizenship*. Cambridge, 2009.
13. J. Conklin et al. Facilitated hypertext for collective sensemaking: 15 years on from gIBIS. In *HYPertext*, 2001.
14. G. Convertino et al. Supporting content and process common ground. In *CHI*, 2009.
15. N. Diakopoulos and M. Naaman. Towards quality discourse in online news comments. In *CSCW*, 2011.
16. M. Dickey-Kurdziolek et al. Lessons from thoughtswap-ing. In *CSCW*, 2010.
17. S. Duncan. The structure of speaker-auditor interaction during speaking turns. *Language in Society*, 3(2):161–180, 1974.
18. S. Faridani et al. Opinion Space: a scalable tool for browsing online comments. In *CHI*, 2010.
19. S. Fussell et al. Coordination of communication. In *CSCW*, 2000.
20. S. Hansen et al. Wikipedia, critical social theory, and the possibility of rational discourse. *The Inf. Soc.*, 25, 2009.
21. W. Hill et al. Edit wear and read wear. In *CHI*, 1992.
22. R. Kraut et al. Listener responsiveness and the coordination of conversation. *J. of Personality and Soc. Psych.*, 43(4), 1982.
23. T. Kriplean et al. Supporting reflective public thought with considerit. In *CSCW*, 2012.
24. C. Lamm et al. The neural substrate of human empathy. *J. of Cog. Neuroscience*, 19(1), 2007.
25. C. Lampe and P. Resnick. Distributed moderation in a large online conversation space. In *CHI*, 2004.
26. K. Nam and M. Ackerman. Arkose: reusing informal information from online discussions. In *GROUP*, 2007.
27. F. Polletta and J. Lee. Rhetoric in public deliberation after 9/11. *Am. Soc. Rev.*, 71(5), 2006.
28. W. Sack. Conversation Map: An interface for very large-scale conversations. *J. Manage. Inf. Syst.*, 17(3), 2001.
29. F. M. Shipman, III and C. C. Marshall. Formality considered harmful. *Comput. Supported Coop. Work*, 8, 1999.
30. M. Smith and A. Fiore. Visualization components for persistent conversations. In *CHI*, 2001.
31. A. Sukumaran et al. Normative influences on thoughtful online participation. In *CHI*, 2011.
32. M. Wattenberg and D. Millen. Conversation thumbnails for large-scale discussions. In *CHI*, 2003.
33. S. White. Backchannels across cultures. *Language in Society*, 18(1), 1989.
34. S. Whittaker. *Theories and methods in mediated communication*, volume 18. New Jersey: Erlbaum, 2003.
35. S. Whittaker et al. The dynamics of mass interaction. In *CSCW*, 1998.
36. V. Yngve. On getting a word in edgewise. In *Regional meeting of the Chicago Ling. Soc.* Chicago Linguistic Society, 1970.