

One and Done: Factors affecting one-time contributors to ad-hoc online communities

Brian J. McInnis¹, Elizabeth L. Murnane¹, Dmitry Epstein², Dan Cosley¹, Gilly Leshed¹

¹Information Science, Cornell University, ²University of Illinois at Chicago

¹{bjm277, elm236, drc44, gl87}@cornell.edu, ²dmitry@uic.edu

ABSTRACT

Often, attention to “community” focuses on motivating core members or helping newcomers become regulars. However, much of the traffic to online communities comes from people who visit only briefly. We hypothesize that their personal characteristics, design elements of the site, and others’ activity all affect the contributions these “one-timers” make. We present the results from an experiment asking Amazon Mechanical Turk (“AMT”) workers to comment on the AMT participation agreement in a discussion forum. One-timers with stronger ties to other Turkers or feelings of trust for Amazon are more likely to leave more — but shorter and less relevant — comments, while those with higher self-efficacy leave longer and more relevant comments. The phrasing of prompts also matters; a general appeal for personally-reflective contributions leads to comments that are less relevant to community discussion topics. Finally, activity matters too; synchronous activity begets responses, while pre-existing content tends to suppress them. These findings suggest design moves that can help communities harness this “long tail” of contribution.

Author Keywords

Online community; One-time participation; Amazon Mechanical Turk

ACM Classification Keywords

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

INTRODUCTION

The “long tail” of participation is a well-established observation in the study of content-oriented online communities [43]. For example, most registered Wikipedia users make only one or a few edits during their first 24 hours of membership and then never return [37]. Q&A forums and discussion boards have similar dynamics, with many members posting exactly one question [20] or comment [53].

These “one-time”, low-volume contributors matter. In RegulationRoom, an online platform designed to help citizens

discuss proposed federal regulations [12], over two thirds of contributors to a recent policy discussion were one-timers, contributing over one third of the comments. In the Stack Overflow family of sites, although frequent users provide many answers, a large fraction of questions come from one-timers [32]. Even in communities with large, active cores, one-timers matter. An informal study by Aaron Swartz suggests that in Wikipedia, substantive contributions to articles often come from infrequent contributors and are then integrated into Wikipedia by core members [50].

One-timers may play an especially important role when the online community is more temporary or ad-hoc. Protest movements, disaster relief efforts, and civic participation forums often arise around a cause and exist for a short time [14, 48], which makes it difficult both to accumulate a core set of members and to develop, teach, and enforce norms. Since time is of the essence to such ad-hoc communities, it is especially important that one-timers get their contribution right.

Generally, these one-time contributors do not spend time lurking on the website and learning the ropes [38], nor do they learn from feedback about their contributions [27] or move from the periphery of the community toward the core [4]. Instead, they make their contribution and move along. Thus, mechanisms such as badges, reputation systems, feedback, and moderation that are commonly used to socialize newcomers, encourage their contributions, and regulate their behaviors [24] are unlikely to help these transient but potentially valuable one-timers. Different strategies are therefore needed to maximize the value that both one-timers and the community might gain from their brief visit to an online space.

To learn more about factors that affect one-timers’ contributions, we conducted an experiment in which Amazon Mechanical Turk workers (“Turkers”) were asked to spend ten minutes testing the interface of a RegulationRoom-like site that presented the Amazon Mechanical Turk (“AMT”) Participation Agreement as the policy under discussion (see Figure 1). We looked at how many comments and words commenters posted, as well as a hand-coded measure of comment *responsiveness* that reflects how closely each comment responds to the norms of a community discussion topic.

Participants’ behavior offers insights about how personal characteristics, design elements, and social activity affect what a person contributes. At the level of personal characteristics, we found that participants with stronger connections to other Turkers and higher trust in AMT made more, but

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CSCW '16, February 27-March 02, 2016, San Francisco, CA, USA
Copyright © 2016 ACM ISBN 978-1-4503-3592-8/16/02 \$15.00.

DOI: <http://dx.doi.org/10.1145/2818048.2820075>

less responsive, comments. By contrast, participants who reported higher self-efficacy made longer and more responsive comments. At the design level, we varied the default “placeholder” text inside the comment box with the goal of guiding one-timer contributions in several ways. However, varying this text did not affect commenting quantity; further, prompts that called for participants to reflect on their experience as Turkers were associated with lower levels of responsiveness. Finally, we measured the effects of social activity by examining comments other Turkers added before and during a participant’s interaction with the site. The presence of more pre-existing comments (added before the Turker entered the site) reduced commenting, while comments that appeared during a Turker’s visit increased it. Further, pre-existing comments with low responsiveness to the policy predicted higher responsiveness, and vice versa.

Below, we first introduce RegulationRoom, the example that motivated our questions and experiment, then discuss the related work on personal, design, and activity factors that influence one-timers’ contributions. We then present our study, its findings, and their design implications around addressing those factors. We conclude with limitations of our study and future research toward understanding and supporting the long tail of one-timers in online communities.

REGULATIONROOM: A MOTIVATING EXAMPLE

RegulationRoom¹ is the online policy deliberation platform [12] that inspired us both to consider one-timers and to design our experiment. Working with federal agencies such as the Department of Transportation, RegulationRoom hosts consultations on federal policy proposals, each lasting a few months. Visitors can comment on proposed regulations, sharing both experiences and knowledge that policymakers can use in the regulation-making process. The project seeks to engage individuals directly affected by the proposed policies, especially those unlikely to participate in existing public consultation processes. For example, in a 2010 consultation around Electronic On-Board Recorders in trucks, the project sought input from individual truck drivers who are traditionally less likely to participate than industry lobbyists.

This goal of including novice participants comes with challenges, including helping people realize they should contribute and translating the “legalese” of federal policies into material that is accessible to a layperson [12]. The design of RegulationRoom addresses these challenges through careful presentation of information, tight ties between agency-provided information and comments, and human facilitation.

First, the information presented is crafted to be accessible to the public and to focus on specific issues. To do this, the RegulationRoom team “translates” proposed regulations to plain English, segments them into topics and subtopics, and connects them to the original policy documents. With agency input, the team also poses specific questions that appear at the bottom of each subtopic to guide contributors as they comment on the proposed policy.

¹<http://www.regulationroom.org>

Second, the site design presents the subtopics as expandable folder webpages, supporting attention to specific aspects of interest while providing context. Subtopic pages have two columns, one with the sub-topic text, the other with a comment thread devoted to that subtopic. The purpose of this two-column design is to orient people toward making comments relevant to specific subtopics. Figure 1 shows a screenshot of the RegulationRoom-inspired interface used in the study.

Third, RegulationRoom relies on a team of graduate (mostly law) students trained in facilitation, conflict resolution, and the substance of the ongoing consultation. Their goals are to elicit effective feedback and engage commenters. Moderators monitor the discussion, welcome newcomers, and respond to comments with guidance about effective participation. This guidance is needed because comments in RegulationRoom are often missing needed information; opinions are often unsupported by facts or personal experiences, or comments do not address the specific issues being presented.

However, for one-timers, who do not return after commenting, this guidance is not helpful because it will never be seen — and most RegulationRoom visitors are one-timers. Approximately two-thirds of visitors are one-timers, and they contribute just over one-third of the total comments; roughly 40% of participants post just one comment. Only one-third of commenters who receive moderator feedback came back to post again. Even in the most popular consultation to date, contributors’ median visit duration was one day, and over one-third of all comments were made by one-timers.

RELATED WORK

The overarching question of our work is: *What factors contribute to valuable one-time participation?* By *valuable* we mean both *more* contributions and contributions that are responsive to the norms of the community and the goals of the discussion. Previous literature has identified a number of factors that might affect one-timers’ contributions, which we group into three broad categories: personal characteristics, design elements, and others’ activity.

Personal Characteristics

Several aspects of a person’s characteristics and background have been shown to affect how people contribute to communities. Key among these is personal interest in the topic of the community, which provides intrinsic motivation to participate [33]. Thus, many communities recruit members with a topical interest in the group [10, 36]. We designed this into our experiment, asking Turkers to comment on a policy of direct interest to them: the AMT Participation Agreement. Turkers were also compensated, providing extrinsic motivation. Since the design encouraged participants to have strong motivation and since the value of intrinsic motivation is well-established, we do not focus on motivation in this study.

Regulation Room
Home

Open Rule

Amazon Mechanical Turk Participation Agreement

Discussion | **Payment for Work** - 11

Select other topics

Subtopics collapse all expand all

1 | When HITS are rejected 4

When HITS are rejected

The mTurk Participation Agreement sets up two options for what happens when a HIT is completed: (1) the Requester approves the work, and mTurk automatically processes a payment to the Turker; or (2) the Requester rejects the work, and the Turker gets nothing. All the Agreement says about rejection is that the Requester can reject 'if the Services do not meet the Requester's reasonable satisfaction.'

Some Turkers worry that Requesters have no accountability for rejecting work on HITS. Rejection not only keeps Turkers from getting paid for their work but also damages their approval rating. Yet Requesters aren't required to respond to Turkers' questions, and Turkers have no way to appealing rejections. From Requesters' point of view, payment is the only leverage they have to insure the quality of work. And, since a Requester is typically requesting and reviewing a large number of HITS, having to answer questions from individual Turkers could become a huge burden.

Should there be clearer standards, or some more formal process, for rejecting HITS?

Should Turkers receive partial payment for rejected HITS?

If so, how would that amount be set?

Comments 4

Make a Comment

Comment Cancel

4:07 PM August 13, 2014

I feel that the requester should communicate more with the turker, if there was a rejection. I agree with the partial payment, I feel there should be some kind of compensation given for the time spent on certain types of hits.

Reply

user@7e4reG1 2
4:06 PM August 13, 2014

There should be partial payment. I had a HIT rejected once because it was 10 words under what the paragraph needed to be. I spent alot of time on it but I assumed I had enough characters but I didnt and I got no payment.

Reply

user@437Ab2 3
4:02 PM August 13, 2014

I believe that Turkers should receive atleast 25% of

2 | How Turkers are paid 2

3 | mTurk's power to delay payments 5

All topics

Turkers' employment status | Payment for Work | Relationships between Turkers and mTurk

Figure 1. User interface screenshot, showing the policy (left) and commenting interface (right) for the first subtopic under the Payment for Work topic. Users can freely move between topics and subtopics using the controls at the top and bottom of the interface. The default text “Make a Comment” is visible in the comment-entry box.

Instead, we look at factors associated with people’s evaluations of the community. In particular, people who feel a sense of connection to other community members or to the group as a whole may be more motivated to contribute. In civic participation contexts, for instance, people who have existing relationships with groups related to a policy area contribute more in discussions [23]. In the context of a discussion by Turkers around the AMT Participation Agreement, this would refer to participants’ connections with other Turkers. Also, people are more likely to be active participants when they perceive the decision-making institution as trustworthy, fair, and legitimate — i.e. have greater institutional trust [51, 18]. In our context, this refers to participants’ trust in AMT to treat them fairly.

Finally, a person’s belief in their ability to make useful contributions matters. Although people may know information or have opinions and experiences related to a policy area, they may feel unable to express these feelings in a policy discussion. This belief in one’s own capability to complete a task or perform successfully is referred to as self-efficacy [1], and perceptions of self-efficacy have been shown to enhance engagement in online political contexts [19]. We therefore hypothesize that one-timers’ personal characteristics around connection to the community of Turkers, trust in AMT, and self-efficacy will influence the number, length, and responsiveness of their contributions.

H1a: Participants who are more strongly connected to other Turkers will be “better contributors” (i.e., make more, longer, and more responsive contributions).

H1b: Participants who have higher trust in AMT will be better contributors.

H1c: Participants with a higher sense of self-efficacy will be better contributors.

Design Elements

A number of studies have shown that site design features can influence newcomers’ initial contributions. For example, removing barriers to participation such as the need to create an account can scaffold initial contributions [4]. The overall feel of the site can also signal norms of contribution. For instance, a friendly atmosphere can encourage newcomers to ask questions [34], while a professional design can lead to posting more thoughtful comments versus an unprofessionally designed site [49]. Even design decisions that seem subtle can affect contribution behavior; making the “edit” tab visible at the top of a wiki page signals that contributions are possible [4], a bigger comment box leads to longer, more thoughtful comments [49], and though pop-up windows can be annoying, they are effective short-term mechanisms to increase commenting [53].

In this study, we focus on the effect of explicit guidance, presented via the placeholder text inside the comment box. Such text is often used as a subtle prompt to instruct users about what kind of information is required (e.g., date, email address, numeric formats). Although the placeholder text may be barely noticed and disappears as soon as a user begins typ-

ing, psychology research suggests that it may prime people’s behavior [2]. In particular, the default text can signal behavioral norms and guide individuals toward desired behaviors [8]. This design element may thus be especially important for one-timers, whose limited experience means they must rely on design cues to understand contribution norms.

In RegulationRoom, the current default text is a *generic* “Make a comment”. This is a clear call to action, but it does not say *how* to take effective action. We hypothesize that default text that explains the *overall goal* of a comment (for example, asking the commenter to read the policy and connect it to their personal experiences) will reduce ambiguity around expectations and increase the amount of contributions. Further, we hypothesize that default text that calls for addressing the *specific* issues being discussed (for example, by asking a specific question related to the policy section being viewed) will elicit comments that are more responsive to these issues.

H2a. Participants who see a default text prompt that describes the overall goals of a comment will make more and longer contributions than those who see a generic prompt.

H2b. Participants who see a default text prompt with a specific request will make comments that are more responsive to the request than those who see a generic prompt.

Social Activity

Online communities have a long tradition in applying principles from social psychology research on social influence (e.g., compliance with social norms [9]) to shape participants’ contribution behaviors. Responses to one’s contributions, for example, may affect the quantity and quality of future contributions [7, 27]. Such feedback may increase (or reduce) one-time contributors’ willingness to return—but it has little impact on the quality of those initial contributions, which are often the only contributions one-timers provide.

Still, people might be influenced when they see others’ contributions before or during their visit, and these contributions are generally made visible by default (though, perhaps filtered by posting time or distributed moderation via others’ votes [28]). Even if one-timers do not lurk prior to commenting, the comments that are visible do provide some insight into the community’s norms, which people often want to comply with [9]. Thus, we see others’ activity as likely to affect both the quantity and responsiveness of one-timers’ contributions.

On the quantity side, there is tension between too little and too much content in an online forum. On one hand, more messages, higher member turnover, and varying levels of member contributions have been shown to predict longer online group lifespans [40]. This suggests that lively groups with higher contribution levels may be especially welcoming to one-timers and signal an expectation of more contributions. On the other hand, too much pre-existing content might deter one-timers from contributing if they feel that they have nothing to add, a common reason people lurk in discussion fora [15, 38]. In synchronous settings, seeing posts appear might encourage responses, but rapid post volumes might make the conversation too hard to follow [46].

In many civic participation communities, the problem is more likely to be not enough content rather than too much. Thus, we expect that seeing others' activity has more benefits than risks and that on balance, both pre-existing and synchronous activity will positively affect how much people contribute.

On the responsiveness side, previous studies have shown that participants post comments similar to what they see on the site — e.g., thoughtful [49], factual [47], or personal [31]. Thus, we predict that participants will mimic the responsiveness of the comments they see posted by others.

H3a: Participants who enter a site with more pre-existing content will contribute more and longer comments.

H3b: Participants who see more synchronously added content will contribute more and longer comments.

H3c: Participants' comments will tend to be more or less responsive when they see more or less responsive comments.

STUDY DESIGN: TURKERS ON AMT POLICY

Our original goal was to study the commenting behavior of one-timers in RegulationRoom itself. However, its visitors are reluctant to spend time answering surveys, plus we did not want to adversely affect an actual policy discussion through design manipulations. Thus, to test these hypotheses we developed a study in which Turkers participated in an online commenting task based on the RegulationRoom design and focused on the AMT Participation Agreement². We chose this population and topic because it created a context where participants would have interest in, experience with, and opinions about the subject matter [25]. Unlike existing HCI/CSCW research that has explicitly asked for a Turker perspective about crowdsourcing to propose seller tools [45] or to discuss crowdwork ethics [44], we designed our study so that Turkers would not be more motivated to accept our "human intelligence task" (HIT) on AMT over another because of the topic.

Materials

The discussion interface, shown in Figure 1, is a near-clone of RegulationRoom. We kept the site organization and design but removed external links and links to other policy discussions. The same team that prepares policy documents for deliberation in RegulationRoom processed the AMT Participation Agreement for the experiment. They first chose the parts of the Participation Agreement that were relevant to Turkers, removing pieces aimed at task requesters. They then summarized parts into plain English and divided them into three topics, each of which had two to three subtopics. For each subtopic they created a set of 2–3 questions that ask for the Turkers' input about that subtopic. In total there were eight subtopics: Turker employment status, ownership of work, rejected HITs, paying for HITs, delayed payments, complying with tax laws, AMT's hands-off approach, and revealing personal information to requesters. An example of a topic, a subtopic, and the questions appear in Figure 1.

²<http://www.mturk.com/mturk/conditionsofuse>

Procedure

We posted the task on Mechanical Turk as a "UI experience" study, framing it that way to avoid priming participants to think explicitly about their commenting behavior. We limited participation to US-based Turkers and used AMT workerIDs to reduce the chance that people would participate multiple times. Participants first completed a consent form and then a pre-experience survey with questions about personal characteristics driven by our hypotheses, including involvement with other Turkers, institutional trust in AMT, and perceptions of self-efficacy. After completing the survey, participants were directed to the currently running instance of the interface — a "room". When the current room reached 10 participants, it was closed and a new room spawned. Participants could only view the comments made by others in their room; having a number of such independent rooms, with Turkers entering earlier and later in the lifespan of the room, allowed us to see how people behaved when seeing a variety of levels of social activity.

Once directed to the room, participants were asked to interact with the interface for 10 minutes. Participants could go to pages corresponding to the policy topics, open subtopics and read them, read comments written by other Turkers in the room, and write comments. We did not explicitly ask participants to leave comments (or, indeed, to do any particular activity) in order to minimize influence on their commenting behavior. After 10 minutes, participants were advanced to a post-experience survey with questions about their interaction with the site, manipulation check questions for the default text, and attention check questions.

We estimated the task would take 20–25 minutes to complete and offered up to \$3 for participation, corresponding to an hourly wage of \$7.20–\$9.00. \$1 was paid immediately after completing the task, and the other \$2 for passing attention checks in the surveys and having reasonable latency for both the discussion and the surveys.

Metrics

Response Variables

We operationalized quantity using both the number and the length of comments. Both the *number of comments* and *log total words* are measures at the participant level. Additionally, we use a hand-coded measure of each comment's *responsiveness* to characterize the relevance of the contribution to the sub-topics.

- **Comments (#):** Total number of comments a participant posted.
- **Log (Words #):** Because the distribution of total words was not normal, we used the log of the total number of words.
- **Responsiveness:** Responsiveness was hand-coded on a six-point hierarchical scale to identify how focused a contribution was in response to a specific subtopic and its discussion questions. Details of the coding scheme and its levels are available in the appendix. To simplify modeling and interpretation — and to align with the need of agency partners in a real RegulationRoom context for contributions to

specifically address questions in the subtopic — we considered comments that explicitly address the questions (level 5) and add new insight (level 6) as “high” responsiveness. Comments that were less related to the discussion questions (levels 1–4) were coded as “low” responsiveness.

Personal Characteristics

The pre-experience survey asked participants questions about their connections with other Turkers (H1a), institutional trust in AMT (H1b), and their self-efficacy (H1c) using five-point-scale items where 1 was low and 5 was high.

- **Turker relationships:** Two items about connecting with other Turkers in real life and connecting with other Turkers through online social networks were averaged to a single measure (Cronbach’s $\alpha = 0.86$).
- **Trust and fairness:** Fifteen items about trusting Amazon Mechanical Turk, perceiving it as treating Turkers fairly, and political efficacy with respect to Amazon Mechanical Turk were averaged to a single measure ($\alpha = 0.94$)³.
- **Self-efficacy:** Eight scale items of generalized self-efficacy and confidence in one’s own abilities and skills [6] were averaged into a single measure ($\alpha = 0.92$)⁴.

Design Elements

Each room was randomly assigned to one of three conditions, which differed only in the call-to-action message that appeared as the default text inside an empty comment box. All participants in the same room saw the same call-to-action message.

- **Control prompt:** “Make a comment”. This is the generic default text that currently appears in the RegulationRoom site (and many others), serving as a control condition.
- **Overall prompt:** “Read this section and tell us how it relates to you”. This prompt is designed to inform participants of the overall goal of civic participation in policy deliberation in RegulationRoom [12]: to make comments that are informed by the portion of the policy under discussion (“Read this section”) and that relate to the commenter’s personal experiences (“tell us how it relates to you”). Our intent was that this would increase commenting by making it easier to understand the expectations of behavior (H2a).
- **Specific prompt:** For each subtopic, a 5–6 word version of one of the questions created for that subtopic (e.g., “What should be Turkers’ employment status?”). The design goal was to direct attention toward a specific question in order to increase the responsiveness of comments (H2b).

³Adapted from other institutional trust studies [3, 5, 18, 21, 42, 51].

⁴We 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 [1]. Further, our task required abilities to which various specific efficacy constructs are relevant (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.

Social Activity

When a participant started the UI experience, they could see comments posted by other participants assigned to the same room. We counted how many comments were posted by others in the room before (*pre-existing*, H3a) and during (*synchronous*, H3b) a participant’s time in the room; these were analyzed at the participant level. At the comment level, we measured the responsiveness of comments that were present in the room at the time each comment was created (H3c).

- **Pre-existing comments total:** The number of comments posted by others before a participant entered the room.
- **Synchronous comments total:** The number of comments added by others while the participant was in the room.
- **Pre-comment low:** The number of low-responsiveness comments present in the room before a given comment.
- **Pre-comment high:** The number of high-responsiveness comments present in the room before a given comment.

Mixed-effects Nesting Factors

Finally, we treat room as a nesting variable for all analyses, and we treat participant and subtopic as nesting variables for analyses of comment responsiveness⁵.

- **Room:** We used room as the primary control variable in the analyses, to account for the effect of being in a group with a set of other participants and seeing their comments.
- **Participant:** For analyses at the comment level (e.g., for responsiveness), we used participant ID as an additional control variable to account for the non-independence of comments by the same participant.
- **Subtopic:** Likewise, the specific subtopic may affect contributions because certain subtopics might be more or less controversial or interesting. Thus, we also treated subtopic as a control variable for analyses at the comment level.

RESULTS

Descriptive Statistics

Table 1 presents a high level descriptive overview of the study and factors. About half of the participants who started the task completed it. Those who completed were on average 36 years of age, and about half identified as female. Although most participants chose to comment (an average of 3.4 times during the session), 66 people who completed the task did not comment. Further, through hand-coding the comments we identified 13 (1.1%) that reflected “test” type comments typical of UI design testing (e.g., Lorem Ipsum); however the vast majority were about the AMT participation agreement, though varied in their responsiveness.

A total of 60 rooms were created: 19 for the control, 20 for the overall, and 21 for the context-specific condition. Due to task abandonment during the pre-survey, an average of about 7.5 of the 10 participants assigned to a room actually completed

⁵Comments and words were aggregated across all subtopics for each participant and analyzed at the participant level, so participant ID and subtopic are not used in the per-participant models of quantity.

Participation	Count
Accepted HITs	549
Completed HITs	329
Commenters (Completed)	263
Completed HIT (Did not comment)	66
Commenters (Did not complete)	62
Total Comments	1092
Total “Test” Comments	13
Total Completer Comments	897
Rooms Created	60
Commenters Who Completed	
	M (SD)
Comments per Completer	3.40 (2.16)
Total Words per Completer	108.61 (78.12)
Mean Responsiveness	3.36 (1.15)
Personal Background	
	M (SD)
Turker Relationships	1.40 (0.48)
Trust and Fairness	2.38 (0.51)
Self-Efficacy	2.29 (0.31)
Social Activity	
	M (SD)
Room Membership	7.45 (2.08)
Synchronous Members	3.33 (2.20)
Pre-existing Comments	8.74 (7.71)
Synchronous Comments	1.61 (2.62)
Pre-comment Low	3.46 (4.06)
Pre-comment High	5.19 (4.79)

Table 1. Descriptive statistics capturing user interactions through the experiment interface, participants’ personal characteristics based on the pre-survey responses, and details about social activity in the room. Numbers of participants are counts; other statistics present the mean (M) and standard deviation (SD).

	Control	Overall	Specific
	M(SD)	M(SD)	M(SD)
Comments (#)	3.1 (1.9)	3.4 (2.2)	3.7 (2.3)
Words (#)	112.6 (80.6)	104.3 (84.4)	108.7 (68.8)
Log (Words #)	3.3 (0.7)	3.1 (0.9)	3.1 (0.8)
Responsiveness	3.5 (1.1)	3.2 (1.2)	3.3 (1.1)

Table 2. The mean (M) and standard deviation (SD) for each response variable (i.e., total comments, log of total words, and responsiveness) in each prompt condition (control, overall, and specific).

the UI experience, with an average of 3 people present in the room at any given time.

A total of 263 participants (93 control, 83 overall, 94 context-specific) completed the HIT and provided at least one comment. Some 62 participants (22 control, 26 overall, 14 context-specific) abandoned the task during the 10 minute period; these participants made 195 comments. We do not include these participants or their comments in our analysis, except in computing social activity measures (since their comments were visible to other commenters). Note that these *abandoners* were not statistically different in their personal characteristics from participants who completed the task.

Table 2 shows the statistics for each response variable for the control, overall, and context-specific design conditions. Differences between design conditions were minimal; however, there were differences in both quantity and responsiveness by subtopic, as shown in Table 3 and Figure 2. More comments

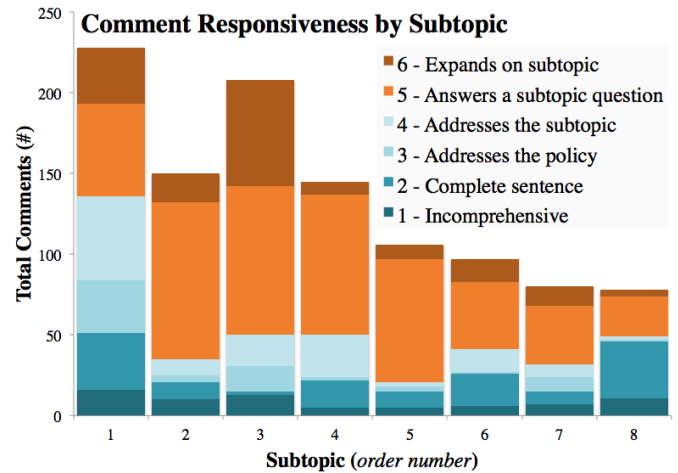


Figure 2. Breakdown of responsiveness by subtopic. Subtopic ID numbers correspond to the ID numbers referenced in Table 3. Levels 1–4 do not directly address subtopic questions and are considered *low* in responsiveness, while levels 5–6 do address and add insight to the questions and are considered *high* in responsiveness.

ID	Subtopic	Comments (#)	α
1	Turker employment status	228	0.85
2	Ownership of work	150	0.823
3	When HITs are rejected	208	0.722
4	How Turkers are paid	145	0.82
5	Power to delay payment	106	0.779
6	Complying with tax laws	97	0.736
7	AMT’s hands-off approach	80	0.774
8	Personal information	78	0.813

Table 3. The number of comments and intercoder reliability for responsiveness (as measured by Krippendorff’s alpha) in each of the subtopics.

were posted to earlier than later subtopics, likely due to order effects. Most comments were rated as high responsiveness, though a larger proportion of comments in subtopics 1 and 8 were rated as low. This may be because Turkers whose primary goal was testing or completing the task would be more likely to do this at the beginning and the end of the task.

Model Interpretation

As the results are discussed in terms of our hypotheses rather than per-model, we first present the models and their interpretation, then present the results (see Table 4).

To predict total comments, log total words, and responsiveness, we built mixed-effects regression models using the personal characteristics, design elements, and social activity variables described above as predictors. For each model, we started with a “beyond optimal” model, incorporating all of the explanatory variables (i.e., personal background, call-to-action message, social activity) as well as interaction terms between them. We then used standard model evaluation criteria including the log-likelihood ratio test, the Akaike Information Criteria (AIC), and the Bayesian Information Criteria (BIC) to reduce the beyond optimal model. All of the interaction terms dropped out of the models because they reduced model performance relative to the model evaluation criteria.

A Poisson mixed-effects regression model was used to predict total comments, as Poisson distributions are appropriate for count variables that have zero probability for negative numbers. As the Poisson is part of the exponential distribution family, the estimate reported for each parameter is its maximum log-likelihood. The Incidence Rate Ratio (IRR) associated with each independent variable can be interpreted as the change in the number of comments when an independent variable increases by one unit and all other independent variables are held constant at their mean value. Thus, an IRR of 1 would indicate that an independent variable had no effect, 0.8 would indicate that there are 0.8 times as many comments for every one unit increase in the specified independent variable, and 1.5 would indicate a 1.5 times increase in comments for each one unit increase in the variable.

A standard linear mixed-effects regression model was used to predict the log of total words, using a normal distribution that fit the log-transformed word counts. We then took the exponent for each of the coefficients to interpret the expected values from the model in terms of word counts as the number of expected additional or fewer words for each one unit increase in the explanatory variable.

Finally, a mixed-effects logistic regression was used to predict high and low responsiveness at the comment level, using a binomial distribution appropriate for binary variables. Its coefficients are interpreted as the expected change that each independent variable contributes to the log-odds (or logits) of a comment being highly responsive. Here we have exponentiated the log-odds to present the odds-ratios. Odds ratios can be interpreted as the change in the response variable expected from a one unit increase to a specific independent variable, holding all others constant.

Personal Characteristics

Our first question was about the relationship between personal characteristics and commenting behavior. Our hypotheses were that stronger prior relationships (H1a), more trust in AMT (H1b), and higher self-efficacy (H1c) would all be associated with producing more comments, longer comments, and more responsive comments.

Our analyses found that prior Turker relationships (H1a) and trust in AMT (H1b) showed a similar pattern. Both were associated with more comments (1.16 and 1.14 times more, respectively), as expected. However, participants with higher trust in AMT contributed fewer words (0.79 times) — opposite what we expected. The relationship to responsiveness was also the reverse of what we had hypothesized for both trust in AMT and Turker relationships (0.68 times for trust and 0.73 times for Turker relationships). It might be that people who trust the system are willing to participate more (hence, more comments) but have less substance to add (hence, shorter and less responsive comments).

Our findings show some support for H1c; self-efficacy was associated with both longer comments (1.59 times) and higher responsiveness (2.37 times) but not with number of comments.

Design Elements

Our second question addressed how manipulation of the default text in the comment box would affect contributions. We hypothesized that the overall-goals prompt would increase the quantity and length of comments (H2a) by giving guidance toward what is expected from comments overall, while specific prompts that ask particular policy questions would increase the responsiveness of comments (H2b).

We were wrong. H2a was unsupported; the overall prompt did not have a significant effect on total comments or on log total words. H2b was also unsupported, with no significant effect of the specific prompt on comment responsiveness. However, an interesting unpredicted effect emerged: the overall prompt elicited lower focused responsiveness (0.36 times).

To better understand this unexpected effect, we examined the responses from the overall prompt condition. We saw comments that provided personal stories about involvement with AMT but that lacked focus on the specific subtopic within the AMT Participation Agreement. The following is an example of such a comment about Turkers' employment status:

"I like the ability to do this as 'extra income' not just as a job. I'm not looking to be hired at pennies on the dollar; I'm just looking to make an extra dollar here and there. I like the time I can work whenever I want too and not at dedicated times. I feel most requesters are fair and treat me with dignity and respect when I have had issues in the past."

Social Activity

Our third question was about the effects of others' activity on the quantity and responsiveness of contributions.

For quantity, we hypothesized that seeing more pre-existing (H3a) and synchronous (H3b) comments would lead to increases in activity. For synchronous comments this was true: each additional comment others posted during a participant's session was associated with a 1.05 times increase in total comments by that participant, as well as a small but significant increase in log total words. For pre-existing comments, however, the opposite was true: for each such comment, participants made 0.98 times as many comments and had a small but significant decrease in log total words.

For responsiveness, we expected that when comments with more or less responsiveness were present, new comments would tend to be more or less responsive accordingly (H3c). Our results suggest the opposite: a one unit increase in pre-existing low-responsiveness comments increased the probability of a high-responsiveness comment by a factor of 1.17, while the presence of pre-existing high-responsiveness comments decreased the odds of a new high-responsiveness comment by a factor of 0.89.

DISCUSSION AND DESIGN IMPLICATIONS

We now discuss how these results might inform the management and design of communities that want to get the most out of their one-time contributors, followed by a discussion of the limitations of the study and questions for future work.

	Total Comments ¹		Log of Total Words ²		Responsiveness ³	
	<i>Std. Error</i>	<i>IRR</i>	<i>Std. Error</i>	<i>Expected Words</i>	<i>Std. Error</i>	<i>Odds Ratio</i>
<i>Design</i>						
(Intercept)	0.26	2.5676 ***	0.39	62.9665 ***	0.74	1.8959
Overall Prompt	0.09	1.0124	0.15	0.8265	0.35	0.3553 **
Specific Prompt	0.09	1.1269	0.14	0.9095	0.33	0.5929
<i>Personal Background</i>						
Turker Relationships	0.07	1.1640 *	0.11	1.0241	0.18	0.7308 .
Trust and Fairness	0.07	1.1413 *	0.09	0.7918 *	0.17	0.6827 *
Self-Efficacy	0.10	0.9137	0.16	1.5882 **	0.27	2.3762 **
<i>Social Activity</i>						
Pre-existing Comments	0.00	0.9765 ***	0.00	0.9994 **	—	—
Synchronous Comments	0.01	1.0535 ***	0.00	1.0013 *	—	—
Pre-existing Low	—	—	—	—	0.05	1.1765 **
Pre-existing High	—	—	—	—	0.04	0.8893 **

Table 4. Mixed effects models predicting total comments, log total words, and responsiveness. 1) Total Comments per participant is estimated with a Poisson mixed-effects model. 2) Log of Total Words per participant is estimated with a mixed-effects linear model. 3) Responsiveness per comment is estimated with a mixed-effects logistic regression. p-value significance codes: ‘*’ 0.001; ‘**’ 0.01; ‘*’ 0.05; ‘.’ 0.1;**

Recruit Effectively with Self-Efficacy

Generalized self-efficacy had a strong, positive association with both the length and responsiveness of comments, as predicted by H1c. The first step in using this finding would be to extend the work on predicting big five personality from social trace data (e.g., [22, 39]) to self-efficacy, trust, and other context-specific characteristics. There is related work that uses self-efficacy to predict social media use (e.g., [26]); that, plus our results that show an association between self-efficacy and commenting behavior, suggest that predicting self-efficacy from online behavior is both plausible and desirable.

Although first-timers will not have generated activity data inside the community that might be used as a basis for predictions, outside activity might be available. For relatively context-independent traits such as personality or generalized self-efficacy, the communities might not even need to be related; personality, for instance, can be modeled with some accuracy based on Facebook “Likes” [22]. Such information might be available to communities that use login credentials from other social media sites such as Facebook, or by linking identities across communities when newcomers choose screen names they also use in other systems [35].

Communities might then choose to recruit only high-efficacy participants to increase the bang for their recruiting buck. This does run the risk of exclusion, though, in a kind of rich-get-richer scenario where only the most comfortable and confident participants contribute. Thus, when recruiting broadly is important (as in many civic participation contexts), these models could be used to identify people who disagree [30] or who have relevant topical interests but low self-efficacy. Developing strategies for recruiting these less likely participants — and helping them develop self-efficacy around contributing — is an open research challenge.

Other personal traits might be important as well. In this study, we saw that social connections to the community (H1a) and trust in the organization (H1b), were associated with more

comments (supporting our hypotheses) but shorter and less responsive contribution (against our hypotheses). One thing we did *not* find in this study was an effect of big five personality factors. Since some studies show that these affect volunteer contributions (e.g., [16, 17]), we had collected personality information as well. However, in models that included self-efficacy, trust, and community connections, none of the personality factors were significant. This might be because evidence of the effect of personality factors is mixed in prior studies and we got “unlucky”, or it might be that factors like incentives or intrinsic interest reduce the effect of personality variables. Our preferred interpretation though, based on the theoretical connections between the characteristics we captured and contribution, is that these are more relevant than the big five for predicting one-timers’ contributions.

Sweat the Small Design Details

Our design of the default prompts did affect people’s behavior—just not in the ways we intended. Contra H2a, the overall prompt did not increase the number or length of comments (and neither did the specific prompt). Contra H2b, the specific prompt did not increase responsiveness, while the overall prompt actually led to lower responsiveness.

Our post-hoc diagnosis is that the prompt designs were aimed at barriers that had been found in other civic participation contexts — but not directly at the criteria underlying the responsiveness metric. For instance, recall that our “Overall Prompt” was intended to both call attention to a particular subtopic in the policy (“Read this section”) and solicit personal experiences with it (“tell us how it relates to you”). The personal experiences portion was driven by observations in prior work on RegulationRoom that personal testimony is useful to policymakers and that it might be easier for people to contribute than information or arguments [12]. Likewise, our “Specific Prompt” condition consisted of short forms of the questions posed at the end of subtopics to help participants address issues of direct interest to discussion owners.

The responsiveness coding scheme developed as we tried to pin down the notion of what it meant to be a better contributor through looking at the actual responses we received. Related to the “issue-related thoughts” metric from Sukumaran et al. [49], the agency partners in RegulationRoom want comments that directly relate to the policy and help answer specific, useful questions. The responsiveness metric that evolved has both of those steps; level 4 refers to the subtopic text, while level 5 answers one of the specific questions posed. Framed in terms of the metric, the overall prompt supports reference to the text — but not to the questions, while a specific prompt supports reference to the question — but makes it possible to answer it without referring to the text. Oops.

Designing with the responsiveness coding scheme in mind might have led to other prompts (e.g., “Read the text and tell us what you think about one of the questions it poses”) that might have provided more guidance, more direct connection to the text, and clearer links to outcome metrics. More generally, these mismatches between design goals and metrics (or between theories of behavior and design implementations [29]) can be pernicious. Other outcome goals would demand their own metrics, and prompts; for instance, in a community that valued engaging with other commenters, the prompt (and other interface elements) would want to emphasize explicit references to another person’s comment.

Further, small details (such as the default text in the comment box) can have strong, unexpected impacts on behavior—a finding that both designers and researchers should be on their toes for. For instance, going back to the example engagement-with-others goal, one might imagine that “explicitly refer to another comment” versus “explicitly refer to another commenter” might lead to quite different effects on the tone of the discussion.

Create (the Right) Amount of Liveliness

Our findings around the effect of others’ activity suggest that liveliness is an important characteristic that affects one-timers’ contributions. This effect is not as strong as that of personal characteristics, but it is there; aligned with H3b, synchronous comments increased both the number and length of comments⁶, while contra H3a the presence of pre-existing content reduced both.

Designers have leeway in how to present and organize others’ activity. Lucky communities that get “too big” could consider running parallel smaller communities, as in this study. This is infeasible if there are strong task-related (synthesis might require seeing all contributions) or ethical (all voices should be heard by all members, as might apply in a civic participation context) reasons to ensure that all content is visible to all members. However, many tasks such as brainstorming and commenting might benefit from parallel discussions where social influence and production blocking are reduced. Imagine a system that uses voting to represent community interest or approval; Sagalnik et al. [41] and Lampe et al. [28]

⁶With the caveat that our discussions were fairly low-volume. Based on the tensions described earlier, in active systems we would hypothesize a quadratic term for liveliness: not too little, not too much, but “just right”.

have shown that initial reactions to items can have long-term impacts on the attention they receive in the system. Using multiple parallel groups might be one way to blunt the impact of these initial conditions.

For the too-small case, one option is to delay the presentation of pre-existing comments, making them appear synchronously. This was used in the ESPGame to support solo play [52]; our results suggest it should increase both the amount and length of commenting. A related idea would be to automatically post canned questions, not in the problem text or in the comment box, but in the form of synchronous posts that appear only to newcomers. These prompting questions, once answered, could be merged with the one-timer’s answer and appear as a single post by that person.

These strategies would probably be ineffective to apply directly to long-term contributors; the effects of strong, explicit calls to action quickly fade [53]. But they did increase commenting in the short term, which is perfect for the use case of one-timers—whose comments might in turn start discussions among more long-term members and benefit the community as a whole.

LIMITATIONS AND FUTURE WORK

We close with a discussion of technical considerations for other people doing related studies, then call out related, interesting questions that our design could not answer but that are fruitful sources of future work.

Technical Issues

We start with smaller choices that might have affected our results. We think the results are robust against these, but we report them for both honesty and potential usefulness for other researchers.

Our choice to assign people to rooms at the point of task acceptance rather than at pre-survey completion meant that in practice our rooms were smaller than our intended sample size of 10, reducing the range of social activity data available to us. Although on average participants experienced the 10 minute discussion environment with 3–4 other people, perhaps the effect of social activity would be stronger with 10 people actively engaged within a room, or perhaps additional people would lead to a feeling of being “crowded” or overwhelmed by activity.

We may have also blunted our ability to see large differences through our choice of a 10 minute window for using the site, right-censoring the ability of more active Turkers to engage with other participants and the material beyond the 10 minutes. We chose 10 minutes to keep the task from being overly burdensome for participants, but it would be useful to validate these results in longer experiments (or more natural contexts).

Browser-related errors are also common in large scale and interactive online experiments. Around four percent of participants were affected by a JavaScript bug that allowed them to comment for more than 10 minutes. Rather than exclude these participants entirely, we chose to only analyze their comments made within the first 10 minutes. Additionally, the average usability rating of the website was only about 2.4 on

a 5-point scale. Ease of use can affect how willing people are to interact with technologies, and this too might have affected our results (including the relatively low completion rate).

Future Work

We now turn to the questions that we could not answer with this design.

How does intrinsic interest affect commenting behavior?

A criticism of using AMT to recruit participants for lab-based social research is that the participants are paid to perform a task. We argued, and believe, that compensating Turkers to consider the Turk policy was an effective way to recruit an informed audience that had both extrinsic and intrinsic motivation to participate [25]. In other words, we tried to minimize the effect of motivation to focus on the constructs described earlier. We felt comfortable about this since the importance of self-interest is a robust finding in the literature, but comparing its effect to other factors would be useful follow-up work.

How could an online community adapt to commenting needs?

Sites can also “ask” the community directly for specific content. For instance, recommending tasks that both match a contributor’s interests and community needs (“intelligent task routing”) can increase people’s contributions in Wikipedia [11]. Task routing itself is less appropriate for one-timers because they may have no interest profile, but related design strategies could help maximize their chance of making a useful contribution. For instance, interfaces could highlight discussion topics that are less well-covered by others [54] or present existing comments that are related to the one being currently written; both strategies could help people avoid excessive duplication, as is often found in Amazon reviews [13]. Another way to help newcomers is to provide explicit guidance about how to make useful comments, for instance through structured templates for contribution [47].

Which one-time visitors are likely to comment?

We also chose to focus only on the behavior of people who actually made at least one comment. There is a different—and useful—question that could be asked around predicting *which* one-time visitors are likely to comment as part of their visit. Our experimental choices of population, topic, and framing mean that we cannot answer that question; that is left for other researchers. We still see value in understanding the effects of these factors on the people who do comment.

Ecological validity and other communities

Finally, our decisions around Turkers, time limits, and the use of AMT all made sense for our interest in one-time visitors to an ad-hoc community and allowed us to collect survey and observational data that would be difficult in a real community situation. However, with experimental control comes experimental costs. The fact that we explicitly recruited participants (whereas in many communities, people find the community in the course of pursuing their own goals around information seeking) and limited participation to 10 minutes (most communities would in fact like one-timers to become repeat offenders) mean that our conditions of initial participation were somewhat different than they would be in most communities.

It is true that the context will affect how people contribute, and follow-up studies in existing communities and natural settings will be useful in refining our understanding of that. Further, as raised above, questions about who decides to comment and who decides to stick around are important ones that we cannot answer with this study. Still, we expect our results to be useful both in understanding what does affect the behavior of one-time contributors who do decide to contribute and in providing tools for designers interested in one-time contributions. The elements that prior work drove us to study—engagement with and trust in the community, belief in one’s own capabilities, design elements that provide clues about norms, and the presence of social activity—are largely independent of the way people come to find the community.

CONCLUSION

Although online communities are often supported by a strong core of active members, many communities — notably ad-hoc ones — rely on a steady stream of one-time contributors. Much existing research about contributions to online communities has focused on the active core, or moving people into it. Here we report on an experiment to elicit more and better contributions from the “long tail” of one-timers.

Our results provide several actionable factors designers might use to affect one-timers’ behavior: conveying an appropriate degree of liveliness, carefully considering outcome metrics in the details of feature design, and (potentially) recruiting members based on important personal characteristics, notably self-efficacy. More broadly, our work calls attention to this long tail and the unsolved mystery of how to design for this large, but transient, population in online communities.

ACKNOWLEDGEMENT

We thank all the Turker participants, whose thoughtful comments inspired additional research. Thanks also to Marie Kurahashi-Sofue, Josselyn Tsai, Cynthia Farina, Mary Newhart, and the Cornell eRulemaking Initiative (CeRI) team for all their help in the project. This work is supported by NSF IIS 0910664, IIS 1422484, IIS 1405634, and HCC 1314778.

REFERENCES

1. Albert Bandura. 1997. *Self-efficacy: The exercise of control*. Macmillan.
2. John A Bargh, Mark Chen, and Lara Burrows. Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology* 71, 2 (1996), 230.
3. Steven L Blader and Tom R Tyler. Testing and extending the group engagement model: linkages between social identity, procedural justice, economic outcomes, and extrarole behavior. *Journal of Applied Psychology* 94, 2 (2009), 445–464.
4. Susan L Bryant, Andrea Forte, and Amy Bruckman. 2005. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*. ACM, 1–10.

5. Gian Vittorio Caprara, Michele Vecchione, Cristina Capanna, and Minou Mebane. Perceived political self-efficacy: Theory, assessment, and applications. *European Journal of Social Psychology* 39, 6 (2009), 1002–1020.
6. Gilad Chen, Stanley M Gully, and Dov Eden. Validation of a new general self-efficacy scale. *Organizational research methods* 4, 1 (2001), 62–83.
7. Justin Cheng, Christian Danescu-Niculescu-Mizil, and Jure Leskovec. 2014. How Community Feedback Shapes User Behavior. In *Proc. ICWSM*. AAAI.
8. Robert B Cialdini. Crafting normative messages to protect the environment. *Current directions in psychological science* 12, 4 (2003), 105–109.
9. Robert B Cialdini, Raymond R Reno, and Carl A Kallgren. A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology* 58, 6 (1990), 1015.
10. E Gil Clary, Mark Snyder, Robert D Ridge, John Copeland, Arthur A Stukas, Julie Haugen, and Peter Miene. Understanding and assessing the motivations of volunteers: a functional approach. *Journal of personality and social psychology* 74, 6 (1998), 1516.
11. Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. 2007. SuggestBot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th international conference on Intelligent user interfaces*. ACM, 32–41.
12. Cynthia R Farina, Dmitry Epstein, Josiah Heidt, and Mary J Newhart. Designing an online civic engagement platform: Balancing more vs. better participation in complex public policymaking. *International Journal of E-Politics (IJEP)* 5, 1 (2014), 16–40.
13. Eric Gilbert and Karrie Karahalios. 2010. Understanding Deja Reviewers. In *Proc. of the ACM 2010 conference on Computer Supported Cooperative Work*. ACM.
14. Libby Hemphill and Andrew Roback. 2014. Tweet Acts: How Constituents Lobby Congress by Twitter. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM.
15. Khe Foon Hew and Noriko Hara. Knowledge sharing in online environments: A qualitative case study. *Journal of the American Society for Information Science and Technology* 58, 14 (2007), 2310–2324.
16. Gary Hsieh, Youyang Hou, Ian Chen, and Khai N Truong. 2013. Welcome!: social and psychological predictors of volunteer socializers in online communities. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 827–838.
17. Tanja Jadin, Timo Gnambs, and Bernad Batinic. Personality traits and knowledge sharing in online communities. *Computers in Human Behavior* 29, 1 (2013), 210–216.
18. M Kent Jennings and Vicki Zeitner. Internet use and civic engagement: A longitudinal analysis. *Public Opinion Quarterly* 67, 3 (2003), 311–334.
19. Gerald Jordan, Megan Pope, Patrick Wallis, and Srividya Iyer. The Relationship Between Openness to Experience and Willingness to Engage in Online Political Participation Is Influenced by News Consumption. *Social Science Computer Review* (2014), 0894439314534590.
20. Elisabeth Joyce and Robert E Kraut. Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication* 11, 3 (2006), 723–747.
21. Dan M Kahan, Paul Slovic, Donald Braman, John Gastil, and Geoffrey L Cohen. Affect, values, and nanotechnology risk perceptions: an experimental investigation. *GWU Legal Studies Research Paper* 261 (2007).
22. Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110, 15 (2013), 5802–5805.
23. Robert E Kraut and Paul Resnick. Encouraging contribution to online communities. *Building successful online communities: Evidence-based social design* (2011), 21–76.
24. Robert E Kraut, Paul Resnick, Sara Kiesler, Moira Burke, Yan Chen, Niki Kittur, Joseph Konstan, Yuqing Ren, and John Riedl. 2012. *Building successful online communities: Evidence-based social design*. MIT Press.
25. Travis Kriplean, Michael Toomin, Jonathan Morgan, Alan Borning, and Andrew Ko. 2012. Is This What You Meant? Promoting Listening on the Web with Reflect. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, Austin, TX.
26. Nicole C Krmer and Stephan Winter. Impression management 2.0: The relationship of self-esteem, extraversion, self-efficacy, and self-presentation within social networking sites. *Journal of Media Psychology: Theories, Methods, and Applications* 20, 3 (2008), 106.
27. Cliff Lampe and Erik Johnston. 2005. Follow the (slash) dot: effects of feedback on new members in an online community. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*. ACM, Sanibel Island, FL, 11–20.
28. 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*. ACM, 543–550.

29. Kimberly Ling, Gerard Beenen, Pamela Ludford, Xiaoqing Wang, Klarissa Chang, Xin Li, Dan Cosley, Dan Frankowski, Loren Terveen, Al Mamunur Rashid, and others. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication* 10, 4 (2005).
30. Pamela J Ludford, Dan Cosley, Dan Frankowski, and Loren Terveen. 2004. Think different: increasing online community participation using uniqueness and group dissimilarity. In *Proc. of the SIGCHI conference on Human factors in computing systems*. ACM, 631–638.
31. Meethu Malu, Nikunj Jethi, and Dan Cosley. 2012. Encouraging personal storytelling by example. In *Proceedings of the 2012 iConference*. ACM, 611–612.
32. Lena Mamykina, Bella Manoim, Manas Mittal, George Hripcsak, and Björn Hartmann. 2011. Design lessons from the fastest Q&A site in the west. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2857–2866.
33. Abraham Harold Maslow, Robert Frager, James Fadiman, Cynthia McReynolds, and Ruth Cox. 1970. *Motivation and personality*. Vol. 2. Harper & Row New York.
34. Jonathan T Morgan, Siko Bouterse, Heather Walls, and Sarah Stierch. 2013. Tea and sympathy: crafting positive new user experiences on wikipedia. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 839–848.
35. Elizabeth L Murnane, Bernhard Haslhofer, and Carl Lagoze. 2013. RESLVE: leveraging user interest to improve entity disambiguation on short text. In *Proceedings of the 22nd international conference on World Wide Web companion*. International World Wide Web Conferences Steering Committee, 1275–1284.
36. Oded Nov. What motivates wikipedians? *Commun. ACM* 50, 11 (2007), 60–64.
37. Katherine Panciera, Aaron Halfaker, and Loren Terveen. 2009. Wikipedians are born, not made: a study of power editors on Wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work*. ACM, Sanibel Island, FL, 51–60.
38. Jenny Preece, Blair Nonnecke, and Dorine Andrews. The top five reasons for lurking: improving community experiences for everyone. *Computers in human behavior* 20, 2 (2004), 201–223.
39. Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. 2011. Our Twitter profiles, our selves: Predicting personality with Twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*. 180–185.
40. Daphne R Raban, Mihai Moldovan, and Quentin Jones. 2010. An empirical study of critical mass and online community survival. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. ACM, 71–80.
41. Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311, 5762 (2006), 854–856.
42. Wolfram Schulz. Political Efficacy and Expected Political Participation among Lower and Upper Secondary Students. A Comparative Analysis with Data from the IEA Civic Education Study. *Online Submission* (2005).
43. Clay Shirky. 2008. *Here comes everybody: The power of organizing without organizations*. Penguin.
44. M. Six Silberman, Lilly Irani, and Joel Ross. Ethics and Tactics of Professional Crowdfork. *XRDS: Crossroads, The ACM Magazine for Students* 17, 2 (2010), 39–43.
45. M. Six Silberman, Joel Ross, Lilly Irani, and Bill Tomlinson. 2010b. Sellers' Problems in Human Computation Markets. In *Proceedings of the SIGKDD-Human Computation (HCOMP)*. ACM.
46. Diane Slaouti and Gary Motteram. Reconstructing practice: Language teacher education and ICT. *Teacher education in CALL* (2006), 81–97.
47. Jacob Solomon and Rick Wash. 2012. Bootstrapping wikis: developing critical mass in a fledgling community by seeding content. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 261–264.
48. Kate Starbird and Leysia Palen. 2011. Voluntweeters™: Self-organizing by Digital Volunteers in Times of Crisis. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. Vancouver, BC, Canada.
49. Abhay Sukumaran, Stephanie Vezich, Melanie McHugh, and Clifford Nass. 2011. Normative influences on thoughtful online participation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3401–3410.
50. Aaron Swartz. 2006. Who Writes Wikipedia? (2006). www.aaronsw.com/weblog/whowriteswikipedia
51. Tom R Tyler. 1990. *Why people obey the law: Procedural justice, legitimacy, and compliance*. New Haven, CT: Yale University Press.
52. Luis Von Ahn and Laura Dabbish. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 319–326.
53. Rick Wash and Cliff Lampe. 2012. The power of the ask in social media. In *Proc. of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 1187–1190.
54. Anbang Xu and Brian Bailey. 2012. What do you think? A case study of benefit, expectation, and interaction in a large online critique community. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, Seattle, Washington.

APPENDIX: RESPONSIVENESS CODING SCHEME

The responsiveness measure began as a six-point hierarchical scale that captures the degree to which a comment responds to the specific sub-topic of the policy it is posted to. At the lowest level were incomprehensible comments, then comprehensible comments that did not clearly address the AMT policy, comments that referred to the AMT participation agreement in general terms, comments that clearly addressed a particular subtopic, comments that clearly addressed one of the questions presented at the end of each subtopic, and finally comments that in their answers contained new sub-topic-relevant information. Comments were coded in ascending order, meaning that a comment that clearly answered one of the questions presented at the end of the subtopic was required to have met all of the preceding conditions, i.e. being coherent, addressing the AMT policy, and address-

ing a particular sub-topic. Table 5 gives more details about the scheme, as well as examples.

Two coders, one graduate and one undergraduate, were trained in the coding scheme until they reached acceptable levels of reliability. The entire corpus was then coded by the undergraduate coder, with 13–33 percent of the comments in each sub-topic also coded by the graduate coder in order to monitor reliability. Monitoring reliability by sub-topic was important because of the content-dependent nature of the coding. The intercoder reliability as measured by Krippendorff's alpha for each sub-topic ranged from 0.72 to 0.85, which is a good level of reliability. All disagreements were resolved through consensus and included in the final analysis. Training data was recoded by the undergraduate coder at the end of the process and included in the final analysis.

Level	Examples
<i>(1) Incomprehensible comment:</i> Comments that are single words or where it is impossible to infer what the commenter refers. Comments that are a pure copy-paste from the policy text.	<ul style="list-style-type: none"> • Independent contractor • By not giving out the SSN of workers! • Yes
<i>(2) The comment is comprehensive, but not clearly about MTK policy:</i> The comment should stand on its own. The comment should be a complete sentence.	<ul style="list-style-type: none"> • I think if you are selecting to do this job it is for the freedom of the restraints that most 9 to 5 jobs don't offer. We knew what we were accepting when we got into it.
<i>(3) The comment refers to MTK policy:</i> The comment has to contain at least some of the terminology specific to the MTK policy. The use of words such as 'Turker', 'requester', or 'HIT' are good signals for inclusion.	<ul style="list-style-type: none"> • I think Turkers should have to complete the tasks in order to get paid. This is not a full time job. • It would be nice to see a more standard amount of time for payments to be processed. 30 days for a dollar is a long time to wait. I understand they need to make sure the data is good, but they should do it in a reasonable amount of time.
<i>(4) The comment relates to a particular sub-topic:</i> The comment has to contain at least some of the exclusive terminology of the sub-topic.	<ul style="list-style-type: none"> • I think that the jobs that we do as Turkers dont really constitute employee status. Not to say it's not important or work, but I feel that Amazon has no priority over what we do [sub-topic 1] • i think the matter has pretty well been gone over - unless you're doing a lot of work for a single requester, you don't have to file; but, if there's going to be a change in this, AMT will be required to do the 1099s anyway so that's not really an issue ... [sub-topic 6]
<i>(5) The comment answers a question posed at the end of the sub-topic:</i> The comment has to contain elements of the question. Also, if a question is about opinion, the answer should contain an opinion; if it is about facts, the answer should contain facts.	<ul style="list-style-type: none"> • I think the joint bank account should be allowed in some circumstances, such as shared spouse accounts. However, I am not sure how they would handle this in a way to keep it from being abused. Other than that, I am completely satisfied with how I get paid. [sub-topic 4] • Most of the payments are getting fast for turkers, but in some cases it is late because the requesters need to get all the responses. AMT can be justified in holding payment because it has the right to do so when a turker do anything against their policies. [sub-topic 5]
<i>(6) The comment contains new sub-topic policy relevant information:</i> The comment has to add new information, not present in the sub-topic. It should introduce a concept not discussed in the policy and provide some background to the answer.	<ul style="list-style-type: none"> • If a Requester is not going to pay for a hit, then the data submitted should not be used. Just like Requester can auto approve payments, workers, should be able auto approve use of their work when payment is made. If no payment is made, the work content is NOT approved for use. [sub-topic 2] • I think there should be some type of appeal button that makes the requester know of this. If they see a large number then they may realize an error has been made and that would cut down on the time from answering 500 messages. [sub-topic 3]

Table 5. A description of the responsiveness coding scheme.