"Guess what! You're the First to See this Event": Increasing Contribution to Online Production Communities

Corey Jackson, Kevin Crowston, Gabriel Mugar, Carsten Østerlund
School of Information Studies
Syracuse University
Syracuse, New York, USA 13210
{cjacks04,crowston,gmugar,costerlu}@syr.edu

ABSTRACT

In this paper, we describe the results of an online field experiment examining the impacts of messaging about task novelty on the volume of volunteers' contributions to an online citizen science project. Encouraging volunteers to provide a little more content as they work is an attractive strategy to increase the community's output. Prior research found that an important motivation for participation in online citizen science is the wonder of being the first person to observe a particular image. To appeal to this motivation, a pop-up message was added to an online citizen science project that alerted volunteers when they were the first to annotate a particular image. Our analysis reveals that new volunteers who saw these messages increased the volume of annotations they contributed. The results of our study suggest an additional strategy to increase the amount of work volunteers contribute to online communities and citizen science projects specifically.

CCS Concepts

• **Human-centered computing** → *Empirical studies in collaborative and social computing;*

Keywords

citizen science; online communities; experiment; motivation; intention to treat; novelty

1. INTRODUCTION

Ciampaglia et al. [3] note that online voluntary production communities have two options to increase the amount of content generated: increasing the number of new volunteers or increasing the participation and retention of existing volunteers. The research presented here is an exercise in the latter: increasing the contribution of existing volunteers. To increase contributions, online voluntary community managers can make motivational aspects of volunteers' participation more salient and thus increase the likelihood of volunteers' contribution. Motivation in online communities has attracted a significant amount of attention in communities like Wikipedia and free/libre open source software (FLOSS) projects. Studies on Wikipedia, for example, have noted the complex web of motivations for volunteers ranging from recognition of contributions by peers [5] to altruism [14].

As well, it is well known that contribution to online communities follows a long-tail distribution (i.e., power law, 80/20 rule) where a handful of volunteers contribute the majority of the content. For instance, in Wikipedia 1% of the editors contribute 55% of the edits[7]. And on the other end of distribution, most volunteers contribute in only one or two sessions and then never return. Because of the large number of such volunteers, encouraging new volunteers to contribute a little more during their "trial period" could be particularly productive for a project.

In this paper, we present the results of an online experiment testing an approach to motivating contribution. Prior research found that citizen science project volunteers described the wonder of being the first person to observe a particular image as a reason for participation in the project. We therefore hypothesized that emphasizing this novelty would increase motivation and thus increase contributions. We implemented a pop-up message that informed volunteers when they were viewing data no other volunteer had seen previously and examined the impact of these messages on the volunteer's level of contribution. The research question guiding this experiment is: How does novelty messaging impact volunteers' contributions to citizen science projects? Though our analysis, we make two important contributions to the literature on motivation in online production communities:

- We show how novelty messaging can be used as a viable solution to increasing contributions to online communities.
- We describe the implementation of an analysis method for experiment research called *intention-to-treat*, an approach commonly used in medical research, but less seen in Group research.

1.1 Citizen Science

Our study is set in the context of an online citizen science project. Citizen science describes projects in which amateur volunteers collect or analyze data to contribute to scientific research. For example, in The Birdhouse Network (TBN), volunteers place nest boxes in their yards and generate and report data they collected from the boxes. Others projects

GROUP '16, November 13-16, 2016, Sanibel Island, FL, USA

DOI: http://dx.doi.org/10.1145/2957276.2957284

work on data (e.g., images, sound files) collected by scientists. In these projects, volunteers perform tasks such as filtering or transcribing data objects. The resulting data is then returned to the scientists to support their investigations. Galaxy Zoo is an example of the latter type of project. Scientists asked volunteers to identify characteristics of galaxies in images collected by the Sloan Digital Sky Survey telescope to support research on galaxy morphology. The success of Galaxy Zoo led its developers to create the Zooniverse [28], a citizen science platform that currently hosts more than forty collaborative projects. The collaboration between professionals and amateurs is technologically mediated through the project website. As with other online voluntary production communities, Zooniverse projects rely on a steady stream of volunteers to complete project tasks.

2. MOTIVATION TO CONTRIBUTE USER-GENERATED CONTENT

How to motivate contributions to online voluntary production communities has received substantial attention from researchers. The link between motivation and contribution is clear: volunteers in the community are motivated; a positive impact on content production is observed. Thus, increasing the motivational affordances of a volunteer's participation is expected to increase the volume of content generated by that volunteer.

The literature on motivation in online communities is quite diverse. Theories such as social loafing, goal-setting, social awareness, social-identity, uses and gratifications, or collective effort have frequently been operationalized in the design of online communities. A major stream of research draws on psychological research on motivation. For example, studies have found that giving volunteers the opportunity to set goals resulted in an increase in contributions to projects [10, 33, 1]. Social motives have also been explored. For example, [30] experimented with making workers in Amazon's Mechanical Turk aware of the presence (social awareness) of other Turkers to increase volunteers' feelings of attachment (bond based and identity based) to the group and found that when Turkers are assigned to work groups and communication between volunteers within work groups is supported, Turkers show more loyalty to the requester and MTurk.

Studies have examined a range of motives in particular settings. For example, in a survey of open-source programmers, Hars and Ou [8] distinguished between external rewards (e.g., peer recognition) and internal rewards (e.g., self-determination) that are the result of participation in OSS environments. Lakhani and Wolf [15] found that one of the strongest motivators for OSS developers is how creative they feel when working on a project, in addition to the intellectual stimulation they get and the utilitarian motivation of improving programming skills. Oreg and Nov [19] found that building reputation and learning were important motives for contribution to FLOSS. Research has identified somewhat different motivations for contribution to the online encyclopedia Wikipedia. Yang et al. [31] discovered that self-concept motivation was the main factor influencing Wikipedians' knowledge-sharing behaviors. Nov [18] found eight different motivations explaining editing behaviors in Wikipedia, including fun, ideology, enhancement, understanding, career, values, protective, and social.

Kraut et al. [13] synthesized the research literature to provide a set of design claims that highlight motivational aspects of volunteers' participation to encourage contribution in online communities. These are grouped in five categories: (1) selection, sorting, highlighting, (2) framing, (3) feedback and rewards, (4) content, task, and activities, and (5) community structure. For example, related to feedback and rewards is the design claim that receiving sincere feedback about performance increases motivation. There is an opportunity then to encourage volunteers to contribute by exploring how social science theories and design claims can be incorporated into the design of online communities.

2.1 Motivation in Citizen Science

Turning to citizen science projects more specifically, researchers have identified a variety of reasons volunteers contribute. In a survey of citizen science volunteers to the Galaxy Zoo project, [21] analyzed more than eighty statements about why volunteers participate in the project. Twelve categories of motivation emerged, including an interest in astronomy (the topic of many of the projects), wanting to help scientists, contributing to science, enjoying beauty, and learning. Reed et al. [23] surveyed 199 users of Zooniverse and identified three factors explaining why volunteers contribute: social engagement with other volunteers, interaction with the site, and helping (or volunteering). The literature on motivation in citizen science projects has also resulted in descriptions of the dynamic nature of motivation suggesting that motivations shift throughout a volunteer's engagement with a project [27, 26].

Many of these motivations are similar to other online production communities like Wikipedia and FLOSS. However, citizen science communities are unique in that volunteers annotate images that few people have seen previously. For example, the Planet Five shows volunteers images that have not yet been seen by anyone outside professional astronomy communities. Indeed, in many projects, images are automatically collected, so they may not have been viewed by anyone. Along with advertising citizen science as a participatory project, many projects note the possibility of viewing such novel images. Research suggests that many volunteers are drawn to projects for this reason. For example, among familiar motives such as interest in astronomy, contribution to science, the beauty of galaxy images, and learning about galaxies, Raddick et al. [21] found volunteers were also motivated by discovery (specifically, "I can look at galaxies that few people have seen before"). Reiss [24] listed sixteen different motives leading to intrinsic feelings. Included in the list was curiosity, described as a desire for knowledge which has the potential to lead to wonder. Jackson et al. [11] found that some volunteers reported being motivated by the possibility to discover new data or find anomalies that others had not previously identified, which is related to novelty.

These motives have support in research that identifies motives for action such as novelty seeking (neophilia), sensation seeking, and curiosity. Addressing consumerism and the desire for novel products, Campbell [2] writes "There are those neophiliacs whose craving for the new takes the form of a preference for the novel, the strange or even the bizarre. These are the volunteers who appear to place a high value on the stimulus which is provided by the unfamiliar while perceiving the known as boring." Raymond [22] introduced the idea of neophilia to describe this trait of hackers: that they are excited and pleased by novelty. Sensation seeking has been researched within the context of internet and technology use. For example, [17] surveyed mobile phone users and found volunteers who scored high on sensation seeking used the phone to make calls more often and tended to use more phone features.

Another possibility is that being first is motivating because it leads to the possibility of discovery, another motivating factor identified in surveys. One of the most well-known cases is the discovery in GalaxyZoo of a novel astrophysical object by a Dutch school teacher, which led to the phenomena being named after her: Hanny's Voorwerp [12]. In another Zooniverse project, Seafloor Explorer, volunteers coalesced around a "stripey tube-dwelling creature" after volunteers asked, "what are the tube shape and the stripey creature?" It was discovered that the images displayed a new species, named convict worm[1] by the citizen scientists. In the Stardust @Home project, when a user discovers a dust grain, he or she is listed as a co-author on the article announcing the discovery and is also given the privilege of naming the dust grain.

There have been efforts to stimulate curiosity as a way to motivate work. In Mechanical Turk, Law et al. [16] showed users obscured (i.e., blacked out) visuals that were only revealed when a worker completed parts of the task. Their research showed the intervention improved worker retention, presumably because they were curious to see the image.

However, to our knowledge, there has been no research on motivation that manipulates perceived novelty as a way to influence volunteer participation. Based on the prior literature on motivation in citizen science and our past research examining motivation in Zooniverse projects, we believe that highlighting the novelty of participation can lead to increased motivation resulting in increased work. We hypothesized that mentioning that a data object has not been viewed by other citizen scientists would appeal to a volunteer's desire for discovery, to be first, or to experience novel occurrences. We therefore offer two hypotheses:

H1: Messaging users about the novelty of their experience increases the number of annotations they contribute.

H2: Newcomers contribute more in their first session when shown novelty messaging than newcomers contributing in their first session not shown a novelty message.

3. SETTING: ZOONIVERSE

The context for this experiment is the online citizen science platform Zooniverse (http://www.zooniverse.org), a web-based system that hosts more than forty science projects. Each project is based around a large collection of data objects (e.g., images, sound/video recordings, text) that require some annotation to support further scientific research. For example, determining the shape of galaxies supports research on galaxy morphology; transcribing information from World War I ship logs supports climate and historical research. Some collections can include hundreds of thousand or even millions of objects to be analyzed, much more than a small science team can handle.

The Zooniverse projects recruit volunteers to do the required annotation. Before analyzing the data objects, volunteers are asked to complete tutorials that explain how to identify relevant characteristics of the data objects, though in most cases, the tutorials are short and can be skipped altogether. Each image is analyzed by multiple volunteers and the individual annotations are compared to determine a consensus answer. At the time of the study, the order of presentation of data objects to volunteers was essentially random, though a few project were experimenting with algorithms for matching data objects to volunteers. In addition to annotating data objects, many volunteers contribute to other aspects of the community, e.g., visiting blogs or posting comments to fora; but these activities are not examined in this study.

[1] http://blog.seafloorexplorer.org/tag/convict-worm/

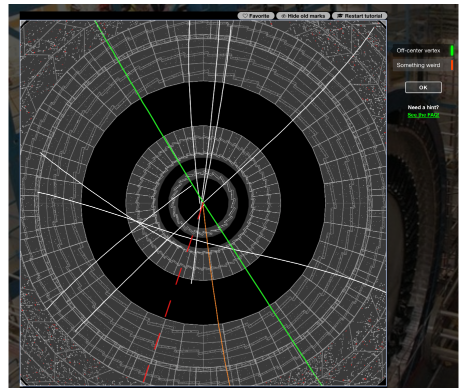

Figure 1: Annotation interface for the Higgs Hunters projects.

3.1 Project: Higgs Hunters

Our study focuses on the Higgs Hunters project. In this project, volunteers annotate images from the Large Hadron Collider (LHC), a particle collider built to search for the Higgs boson particle. The data objects in this project are images of two beams of particles colliding and creating a shower of new particles, possibly including previously unknown particles, such as the Higgs. Charged particles leave a trace in the image; uncharged particles are invisible. The task for the volunteers is to annotate the images for decay anomalies or appearances of off-center vertexes, which are indications that a new uncharged particle was created but then decayed into other charged particles (e.g., the Higgs decays only 10^{22} seconds after it is created). Not all images have off-center vertexes from particles, so seeing a new particle is akin to finding a needle in a haystack.

The annotation interface is shown in Figure 1. To record a find, volunteers click on "Off-centre vertex" on the right-hand side of the window, then mark the location of the vertex and how many tracks appeared. Volunteers are not given any feedback about whether their annotations are correct or if the image is useful for science, because at the time of annotation, neither of those are known. As well, annotations are done independently for the purpose of preventing one volunteer's decisions about an image to influence the annotations done by others.

4. METHODS

4.1 Experiment Design

To manipulate motivation by appealing to volunteers' interest in novelty, beginning in October 2014, the Zooniverse introduced a system that alerted volunteers when they viewed an image that no previous volunteer had seen, specifically, a pop-up message that read "Guess what! You're the first to see this event." The messages were truthful, appearing only for images that had not been previously annotated by any volunteer, so only the one volunteer who was the first to annotate the image saw the message for that image. At the time, the Zooniverse platform did not support random assignment of volunteers to different treatments, so this intervention was added for all volunteers.

Because of this limitation, we studied the impact of the messages using a quasi-experimental design. Project scientists periodically inject new images into the project, resulting in periods where many volunteers see the pop-up message. We used work done during such periods to create the treatment group. However, eventually every image has been viewed at least once, and so the pop-up message is not shown for an extended time as additional annotations are added to the existing images. We used the work done in such periods to create a control group. However, rather than comparing work done by volunteers randomly assigned to treatment or control (a true experiment) we are comparing work done at different times with and without the treatment (a quasi-experiment).

4.2 Data Collection

We collected data from the log files on Zooniverse servers in October 2015. The dataset contained all the annotations done by volunteers up to that time, including a timestamp for each annotation and whether the volunteer was shown the pop-up message (i.e., if they were the first person to see that image). We did not include annotations done by volunteers who were not logged into the system.

The total dataset includes 683,970 annotations contributed by 6,354 volunteers, though not all of these data were used in the analysis. Analysis was done at the session level. Annotations were grouped by volunteer and then into sessions, a series of annotations done by a single volunteer that are separated by a gap of less than 30 minutes. The intuition is that a volunteer generally does some number of annotations in a single sitting (possibly with a short break between annotations) and then takes a longer break, e.g., until the next day. We used a session as the unit of analysis to test the hypothesis that a message might increase interest and so lead to the volunteer's extending the time spent and work done while on the system, resulting in a longer session. There were a total of 17,353 sessions, ranging in length from 1 to 1,504 annotations, with an average of 39 and median of 16 annotations.

For each session, we recorded how many pop-up messages were shown. A total of 28,577 messages were displayed in 3,096 sessions to 1,867 volunteers (i.e., about 18% of sessions had a message and about 29% of volunteers saw a message). Figure 2 shows the number of sessions done per day over time; sessions with at least one message are in blue and sessions with no messages in orange. From the blue areas in the figure it is easy to identify the dates on which new images were added to the system. Immediately after new images are added, a volunteer may see many new objects in a single session: the maximum number of messages in a single session was 419. At other times though, all images will have have had at least one annotation, and a session will have no messages. Note also the very large spike in work done in the first few weeks of the project and the decline in activity over time.

4.3 Within-subjects analysis

We carried out two different analyses on the sessions. Our first analysis was within-subjects, comparing for the same volunteer, the length of sessions that had or did not have a pop-up message. A simple t-test for the difference in session length would be confounded by the large differences in contributions from different volunteers. It might be the case that volunteers who contribute more, e.g., because of higher interest, also see messages more often. We therefore controlled for the individual volunteer by carrying out a repeated measures analysis. Comparing sessions within a subject helped control for the very high variability of contribution to the project, but only partly. Thus, we chose to analyze the data using mixed models since they are more are more sensitive to within-subject variance [4], which is a characteristic of our data. Data were analyzed using the R Statistical package nlme [20]. Since obtaining p-values for mixed models isn't straightforward, our approach relied on comparison of the mixed models using likelihood ratio test which is the probability of seeing the data we collected given a model [**?**]. We compare a null model (i.e, a model disregarding the fixed effect of treatment) and a "full model" with treatment as a fixed effect. To compare the models we use an ANOVA, which yields a chi-squared value, degrees of freedom, and the p-value. With this, we can conclude that the fixed effect of treatment is significant if the difference between the likelihood of these two models is also significant. The advantage of the within-subject design is that it uses more of the data. A possible confound to this design is that at certain points in the project, seeing a pop-up message becomes a matter of chance, as a part but not all of the images are new. As a result, during those periods, rather than messages causing sessions to be longer (our primary hypothesis), a longer sessions increase the chance of novelty messages being shown.

4.4 Between-subjects analysis

To avoid the confound noted above, we also carried out a between-subjects analysis on a subset of the data using a quasi-experimental design. To form the treatment group (sessions with at least one pop-up message), we identified a two-week period during which new images were available (the middle light blue bar in the inset in Figure 2). We chose a period later in the life of the project when the number of sessions done per day was beginning to plateau, meaning that the sample is not from early joiners whose behavior may differ from other volunteers, and the plateau could not be attributed to non-project events such as holidays. To eliminate possible influences from prior experiences on the system, we only used sessions from volunteers who had their first session during the period, meaning this analysis is restricted to the impact on newcomers. For this analysis, we included all sessions done by those volunteers in the two weeks after their first session (i.e., the group includes some sessions beyond the treatment period for volunteers who joined late in the period). Note though that not all sessions during this period had a pop-up message (as shown by the orange part of the graph), an issue we discuss below.

The control group (sessions with no pop-up message) was formed in a similar fashion. However, because the number

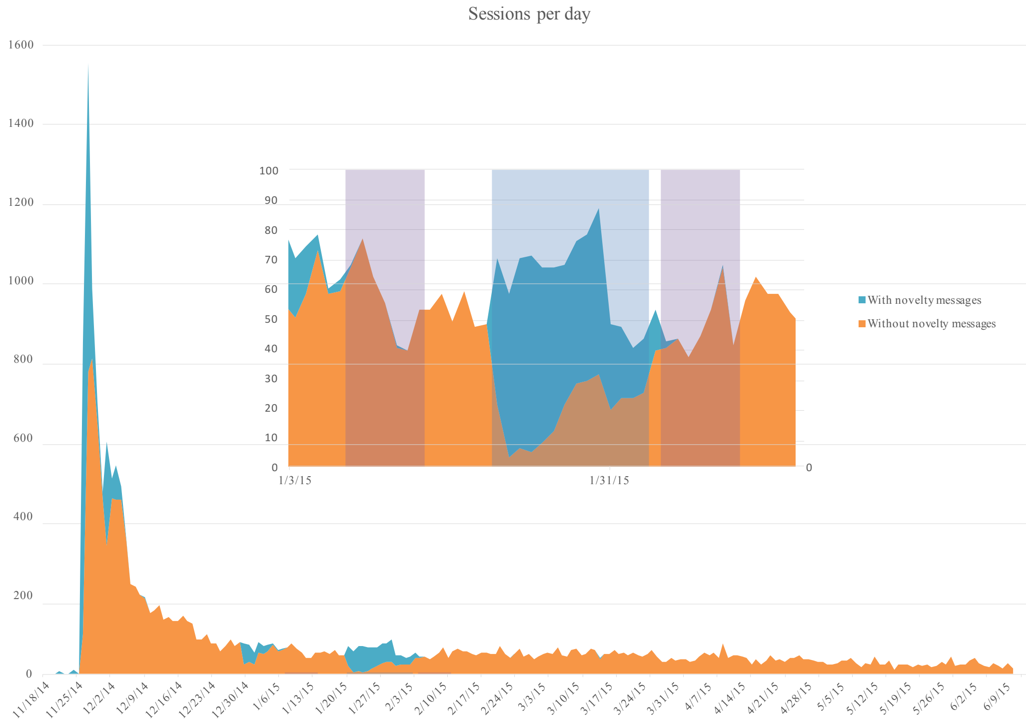

Figure 2: The graph shows the number of sessions per day over time. Sessions with at least one pop-up message are in blue; those without, in orange. The inset shows the periods selected as the treatment period (blue bar) and the two periods selected for the control periods (two purple bars).

of sessions per day was steadily declining over the life of the project, we controlled for maturation by selecting one week just before and one week just after the treatment period. The purple bar in the inset in Figure 2 indicates the first control period, and the last bar, the second control period. Because we followed volunteers for two weeks after their first session, we left a week's gap between the first control period and the treatment period. Even so, we had to drop two sessions from the control group that edged into the treatment period and included a pop-up message. Both control and treatment periods were multiples of 7 days to include all days of the week equally.

5. RESULTS

The contributions of volunteers in online citizen science projects typically follow a long-tail distribution: many volunteers contribute little content, e.g., only a single session, while a dedicated handful of volunteers contribute the majority of content. Higgs Hunters volunteers are no different: most volunteers (71%) contribute in only one session. A handful of dedicated volunteers (eight) contributed more than 11,000 annotations each—a stark contrast in behavior. The average number of sessions by volunteers is 2.73. We computed a 1 percent trimmed mean to illustrate how little work most volunteers contribute. In the trimmed distribution, the average number of sessions dropped to 1.32 (SD=11.52) and the average number of annotations dropped substantially, from 107.65 annotations to 30.47 (SD=699.36). Again pointing to the extremes in volunteers' contribution patterns.

5.1 H1:Pop-ups Increase Contribution

5.1.1 Within-subjects analysis

We analyzed the effects of a message on the number of annotations submitted during sessions. Since the within-subjects comparison relies on a minimum of two observations, we created a subset of cases where a volunteer had at least one session where they saw a message and one session where they did not receive a novelty message. This resulted in 365 unique volunteers and 6,973 volunteer sessions. Novelty messages were shown during 1,355 (19%) sessions. The results revealed that in sessions where messages were shown, volunteers contributed more annotations (see Figure 3). The sessions in which messages were shown had an average of 77.5 (SD=107.6) annotations and a 5% trimmed mean of 55.4 annotations while sessions with no messages shown included on average 41.6 annotations (SD=77.8) and had a 5% trimmed mean of 25.5 - an increase in 35.9 annotations during treatment sessions.

To address the skew in the distribution of session lengths, we log transformed the dependent variable - annotations. Since our data is a repeated measures, we used a linear mixed effects model to compare the number of annotations submitted during sessions of individuals in which they receive a novelty message to those when they failed to receive a message.

To test the significance of this difference, we compared two models: a full model (with the fixed effect treatment) against a reduced model without the fixed effect. As mentioned in the methods, these models can be evaluated using

Experiment Group	Start-End Date (No. days)	No. Volunteers	No. Sessions	Average No. Annotations
Control Prior (CPrG)	1/8/15-1/14/15 (7 days)	107	144	18.1 (SD=26.1)
Control Post (CPG)	2/5/15-2/11/15 (7 days)	76	141	36.8 (SD=53.6)
Pooled Control (PCG)	**14 days**	**183**	**285**	**27.3 (SD=43)**
Treatment (TG)	**1/21/15-2/3/15 (14 days)**	**217**	**356**	**46.4 (SD= 74.7)**

Table 1: Descriptive statistics of treatment and control groups.

Experiment Group		No. Sessions	Average No. Annotations
Treatment (T)	Treated (TT)	223 (62.6%)*	64.7 (SD=84.6)*
	Not treated (T-NT)	133 (37.4%)*	15.7 (SD =37.9)*
Pooled Control (PC)	Hypothetically treated (C-T)	179 (62.6% of 285)**	34.4 (SD =44.4)**
	Hypothetically Not treated (C-NT)	106 (37.4% of 285)**	15.7 (SD =37.9) ***

Table 2: Descriptive statistics for treatment and hypothetically derived control groups. In the table * indicates observed values from population, ** indicates computed values based on observed PC and assumed PC-NT, and * assumed to be same as T-NT.**

Experiment Group		No. Sessions	Average No. Annotations
TreatmentNewcomer (TN)	Treated (TTN-Newcomer)	135	51.9 (SD=69.3)
	Not-treated (T-NT-Newcomer)	82	7.1 (SD=10.9)
PooledControl (PC-Newcomer)	Hypothesized - Treated (HPC-T-Newcomer)	114	24.9 (SD=34.4)
	Hypothesized - Not treated (HPC-NT-Newcomer)	69	7.1 (SD=10.9)

Table 3: Newcomer experiment groups with outcome variable annotations. In the table * indicates observed values from population, ** indicates computed values based on hypothesized control group, and * assumed to be same as T-NT.**

the likelihood ratio test through a chi-squared comparison to determine whether the fixed effect had an impact on the independent variable (i.e., annotations). The results of the comparison revealed that the treatment fixed effect was significant $\chi^2(2)$=171.9, p<0.0001.

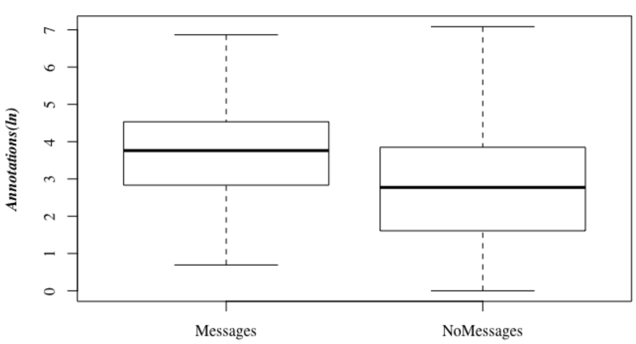

Figure 3: Graph depicting the group differences in the number of annotations in sessions where participants received the treatment and the control.

5.1.2 *Between-subjects analysis*

In the between-subject analysis, we compared the length of sessions done by those who joined during the treatment period with many pop-up messages (the blue bar in Figure 2 to those who joined during a control period without such messages (the purple bars). During the treatment period, 217 new volunteers joined and contributed across 356 sessions (TG). During the first control period, 107 new volunteers joined the project and contributed annotations in 144 sessions, and in the second period, 76 new volunteers joined and contributed in 141 sessions, for a combined total of 183 new volunteers contributing in 285 sessions during the pooled control periods (PCG). The average number of annotations done in those sessions is shown in the final column of Table 1. A t-test shows that the difference between the average session's length in the TG (46.4) and PCG (27.3) is statistically significant (t(639)=3.83, p < .001), suggesting that the pop-up messages do increase the length of a session.

The estimate above is conservative, since as noted above, not all sessions in the treatment group actually experienced the treatment. Specifically, of the 356 sessions in TG, only 223 had a pop-up message (62.6%); the remaining 133 (37.4%) did not. As a result, our estimate of the impact of the treatment is diluted by the sessions in which the treatment was not seen. The 223 sessions in the subset of sessions that were treated (TT) had 64.7 (SD=84.6) annotations on average, while volunteers' sessions in the treatment but not treated (T-NT) subset contributed only an average of 15.7 (SD=37.94) annotations.

To get a better estimate of the impact of a pop-up message, we apply an intention to treat analysis. Figure 4 shows how the subjects in the analysis are divided into control and treatment group and how the treatment group is further divided into treated and untreated groups. We can observe the average length of a treated session (T-T, 64.7 annotations), so the problem is to find a suitable comparison group of untreated sessions. We cannot simply compare T-T to T-NT

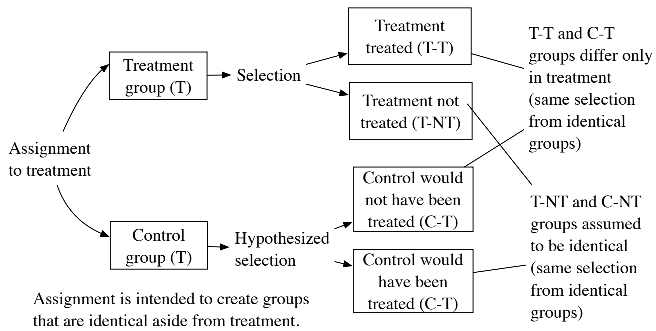

Figure 4: Logic of intention to treat analysis. Regular font indicates observed groups; italics indicate hypothesized groups.

or the entire control group (PCG) because of the confound noted above: at some times seeing a pop-up message is a matter of chance, so a longer session (i.e., from a more interested volunteer) is more likely to have a pop-up. As a result, volunteership in the T-T group is not random, but is related to the outcome variable, meaning the difference between T-T and the other groups could be due to selection rather than the treatment (as shown in Figure 3).

To create a comparison group for T-T we need to select a comparable subset of the control group. Fortunately, it is not necessary to actually carry out the selection; instead, we can do it hypothetically and compute the results. We assume that the control and treatment groups are identical aside from the treatment. Such comparability is the goal of experimental design and is an assumption of a quasi-experimental design. Therefore, had the control group been treated, it would have split in the same proportion as T into a subset that would have received the treatment (C-T, e.g., sessions from more interested control group volunteers) and a subset that would not have received the treatment (C-NT, e.g., sessions from less interested volunteers). As the volunteers of the C-T subgroup are selected in the same way as the T-T subgroup, they should be comparable to the T-T-subgroup, aside from the treatment; and similarly for the C-NT and T-NT subgroups.

We can compute the properties of the C-T subgroup indirectly. Since they are identically selected subsets of assumed-to-be identical groups, T-NT and C-NT are assumed to have identical properties: the same mean number of annotations (15.7) and same standard deviation (37.9). Given the observed properties of C as a whole and the assumed properties of the C-NT subset (the same as T-NT), we can compute the properties of the C-T subgroup to compare to T-T. The results are shown in Table 2. A t-test comparing the mean number of annotations in T-T (64.7, SD= 84.7) to the hypothesized number in C-T (34.3, SD= 44.4) shows a statistically significant difference, $t(400)=4.35$, $p < .001$. As expected, this difference is larger than the difference between the means for the entire control and treatment groups.

Note that these two analyses answer slightly different research questions. The comparison of the whole treatment and control groups shows the impact of implementing the intervention on the expected average length of a session, diluted because not all sessions are treated. The second shows the expected impact of the treatment on a session that actually receives it.

5.2 H2: Intention to Treat - Newcomers

We carried out the same intention to treat analysis examining just the volunteers' first sessions. This was done to examine whether our experimental manipulation had an impact on a population who was unlikely to contribute in future sessions. Such a comparison is interesting because as noted earlier, many volunteers contribute to only one session, so increasing the length of this session may have a big impact on the project. This comparison included 135 sessions in the treatment treated subset (TTN-Newcomer), 82 in the treatment not-treated (T-NT-Newcomer), 114 in the hypothesized pooled control who would have been treated (HPC-T-Newcomer), and 69 in the hypothesized pooled control which would not have been treated (HPC-NT-Newcomer). The descriptive statistics for the population of initial sessions are shown in Table 3. Treated newcomers (TTN-Newcomers) contributed 51.9 annotations (SD=69.3) while those in the hypothesized control (those who would have been treated, HPC-T-Newcomer) contributed only 24.9 (SD=34.4) annotations, a statistically significant difference of 27 annotations, $t(247)=3.8$, $p<.001$.

6. DISCUSSION

Our analysis confirmed both of our hypotheses. First, we found that novelty messaging increases contributions. We showed that message pop-ups alerting volunteers to the fact that they had seen images that had not been viewed by other volunteers almost doubled the number of annotations submitted in a session: from 34.4 to 64.7. We also showed that the messages led to lengthier sessions for the same volunteers (within-subjects): in sessions where volunteers saw a message, they contributed significantly more annotations (77.5 vs. 41.6). Second, we showed that the population most likely to leave the project—namely newcomers in their first session—can be motivated to contribute more prior to their departure. In volunteers' first sessions our experimental manipulation increased the number of annotations from 24.9 to 51.9, again doubling the number of annotations contributed.

Prior surveys and interviews had suggested that volunteers are motivated by seeing images that have not been seen before. However, this paper makes an important contribution by rigorously testing whether making novel experiences known to users works as a method to increase motivation and contribution[2].

6.1 Novelty as a Motivator

Given that very little research in online communities has investigated novelty, we wondered what makes novelty a salient motivator in this context. In Zooniverse it is assumed the citizen scientists are attracted to projects because of the underlying science which supports discovery. Research in psychology offers some explanation. Generally, it is assumed that among other responses, novelty acts as a stimuli to increase interaction and attention of subjects. In a study on child development, Smock and Holt [?] supported a hypothesis which showed that novel objects resulted in significantly more responses (i.e., interaction) than non-novel

[2] As Mark Twain put it, "Supposing is good. Finding out is better."

objects. Berlyne [?] proposed a theory of curiosity motivation which posits that novel stimuli induces longer visual exploration of the stimuli since individuals are curious about the new stimuli. While not frequently cited in the literature on online communities, the introduction of novel data objects (new images in our context) invokes increased attention and curiosity for individuals engaged in the activity. We suspect other communities might benefit from novelty cues to increase individuals' attention and curiosity.

6.2 Implications for Design

The literature on motivation in online production communities like the ones researched in this paper point to a variety of strategies to motivate volunteers to contribute to projects. Our results have clear implications for system designers, specifically that they might want to: a) include such a message if possible and b) spread novel experiences across users and sessions to maximize their impact, rather than having one lucky user see many of them. The most apparent uses of our pop-up might be in citizen science where the goal of the system supports motivation through novel data objects. In the system we studied, there were a total of 28,577 pop-ups (i.e., novel data objects) and only 17,353 user sessions, so there are enough data that each user could see something novel every time they contribute (though implementing such an change might delay processing the data).

We can also imagine how novelty might appear in other communities such as Wikipedia, open source software projects, Q&A communities or blogging sites. In Q& A communities, being the first to respond to a post holds the promise of increased attention to one's comment, i.e., that others after the initial poster will see their responses. The first mover also sets the topic of conversation or is perhaps the first individual to point out a novel feature. For example, in a study of Answerbag, a Q&A site, Gazan et al. [6] found the first submitted answers accumulate 17 percent more rating points than subsequent responses. Gazan et al. [6] also noted: "If there is a first-mover advantage in a social Q&A environment, there must be a measurable benefit to having the equivalent of a dominant market position, regarding some desirable limited good." In online communities where answers are valued highly, being the first to post might make a comment more prominent to readers; in communities where social voting is a feature, it might increase the number of up-votes.

System managers could draw on this motivation in different ways. For example, in open source software communities, designers of projects might highlight novel coding challenges and encourage potential contributors to be the first to solve a particular problem. Wordpress pages encourage authors to publish posts that include a "Be the first to comment" script at the end, in order to encourage readers to start a conversation.

6.3 Encouragement Designs for Similar Experiments

A methodological contribution of this research was the use of an intention to treat approach to analyzing the data (see [9] for more details). The approach was necessary because not all of the sessions selected for the treatment group actually received the treatment. We believe that other interventions in the working of online communities may face the same limitation, meaning that this approach to analyzing the data may be generally useful.

7. LIMITATIONS

There are four limitations we wish to alert the reader to. First, is the experimental design itself: a quasi-experiment. Given the wholesale introduction of the treatment to the population, this design was the only feasible way to analyze the data from the system. A true randomized controlled experiment would have been preferred, but the system was not able to support randomized assignment.

Second, there is the possibility that the effects of the treatment last longer than the single session we analyzed. For example, the effects of a message seen during a volunteer's first session might prompt the volunteer to return for a second session or have continuing impacts in that session. Third, and relatedly, we did not examine whether seeing multiple messages in a session increases the impact of the treatment or if they instead become annoying and a distraction. Future work can explore the temporal dimensions of the effect.

Fourth, while an experiment provides strong evidence for a causal relationship between a treatment and an outcome, there is a trade-off for the richness of data. For example, our analysis does not include possible differences in volunteers that are not captured by the system (e.g., demographics or education). Nor does it provide the kind of rich data needed to illuminate the mechanism of the effect, that is, why volunteers found novelty to be motivating. Prior research found that volunteers report being so motivated, but understanding in more detail exactly why is a question for further in-depth investigation. More research is needed to conclude specifically what drives the observed phenomenon: novelty seeking, discovery, curiosity, sensation seeking, neophilia, some other term or maybe some combination of existing drivers.

8. ACKNOWLEDGEMENTS

Without the Zooniverse volunteers working on the projects and taking the time to answer our questions there would be no paper. Many thanks to the Zooniverse team implementing the experiment. This material is based upon work supported by the National Science Foundation under Grant No. 12-11071.

9. REFERENCES

[1] M. Burke and B. Settles. Plugged in to the community: social motivators in online goal-setting groups. In *Proceedings of the 5th International Conference on Communities and Technologies*, New York, New York, USA, 2011. ACM Press.

[2] C. Campbell. The Desire for the New. In D. Miller, editor, *Consumption Theory and issues in the study of consumption*, page 452. Taylor & Francis, Jan. 2001.

[3] G. L. Ciampaglia and D. Taraborelli. MoodBar. In *CSCW '15: Proceedings of the ACM 2015 conference on Computer Supported Cooperative Work*, pages 734–742, New York, New York, USA, 2015. ACM Press.

[4] A. P. Field, J. Miles, and Z. Field. *Discovering statistics using R.* London ; Thousand Oaks, Calif. : Sage, 2012.

[5] A. Forte and A. S. Bruckman. *From Wikipedia to the classroom: exploring online publication and learning.* International Society of the Learning Sciences, June 2006.

[6] R. Gazan. First-Mover Advantage in a Social Q&A Community. In *2015 48th Hawaii International Conference on System Sciences (HICSS)*, pages 1616–1623. IEEE, Dec. 2014.

[7] R. S. Geiger and A. Halfaker. Using edit sessions to measure participation in wikipedia. In *CSCW '13: Proceedings of the 2013 conference on Computer supported cooperative work*, page 861, New York, New York, USA, Feb. 2013. ACM Request Permissions.

[8] A. Hars and S. Ou. Working for Free? Motivations for Participating in Open-Source Projects. *International Journal of Electronic Commerce*, 6(3), Apr. 2002.

[9] S. Hollis and F. Campbell. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*, 319(7211):670–674, 1999.

[10] C. Jackson, K. Crowston, G. Mugar, and C. Østerlund. Encouraging Work in Citizen Science: Experiments in Goal Setting and Anchoring. In *CSCW Companion'16*, San Francisco, 2015.

[11] C. B. Jackson, C. Østerlund, G. Mugar, K. D. Hassman, and K. Crowston. Motivations for Sustained Participation in Crowdsourcing: Case Studies of Citizen Science on the Role of Talk. In *2015 48th Hawaii International Conference on System Sciences (HICSS)*, pages 1624–1634. IEEE, Dec. 2014.

[12] W. C. Keel, C. J. Lintott, K. Schawinski, V. N. Bennert, D. Thomas, A. Manning, S. D. Chojnowski, H. van Arkel, and S. Lynn. The History and Environment of a Faded Quasar: Hubble Space Telescope observations of Hanny's Voorwerp and IC 2497. *The Astronomical Journal*, 144(2):66–39, July 2012.

[13] R. E. Kraut, P. Resnick, S. Kiesler, Y. Ren, Y. Chen, M. Burke, N. Kittur, J. Riedl, and J. Konstan. *Building Successful Online Communities: Evidence-Based Social Design.* The MIT Press, 2012.

[14] S. Kuznetsov. Motivations of contributors to Wikipedia. *ACM SIGCAS computers and society*, 2006.

[15] K. Lakhani and B. J. Wolf. Why hackers do what they do: understanding motivation and effort in free/open source software projects. In J. Feller, B. Fitzgerald, S. Hissam, and K. R. Lakhani, editors, *Perspectives on Free and Open Source Software.* 2005.

[16] E. Law, M. Yin, K. J. Goh, M. Terry, and K. Z. Gajos. Curiosity Killed the Cat, but Makes Crowdwork Better. In *CHI '16: Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 2016.

[17] L. Leung. Leisure boredom, sensation seeking, self-esteem, and addiction. *Mediated interpersonal communication*, 2008.

[18] O. Nov. What Motivates Wikipedians? *Communications of the ACM*, 50(11):60–64, 2007.

[19] S. Oreg and O. Nov. Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Computers in Human Behavior*, 24(5):2055–2073, 2008.

[20] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2016. R package version 3.1-124.

[21] J. Raddick, G. Bracey, P. L. Gay, C. J. Lintott, P. Murray, K. Schawinski, A. S. Szalay, and J. Vandenberg. Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. *Astronomy Education Review*, 9(1):1–18, Dec. 2010.

[22] E. S. Raymond. *The New Hacker's Dictionary.* MIT Press, 1996.

[23] J. Reed, J. Raddick, A. Lardner, and K. Carney. An Exploratory Factor Analysis of Motivations for Participating in Zooniverse, a Collection of Virtual Citizen Science Projects. In *2013 46th Hawaii International Conference on System Sciences (HICSS)*, pages 610–619. IEEE, Dec. 2012.

[24] S. Reiss. Multifaceted Nature of Intrinsic Motivation: The Theory of 16 Basic Desires. *Review of General Psychology*, 8(3):179–193, 2004.

[25] C. Rhodes and A. Pullen. Editorial: Neophilia and organization. *Culture and Organization*, 16(1):1–6, Mar. 2010.

[26] D. Rotman, J. Hammock, J. Preece, C. L. Boston, D. L. Hansen, A. Bowser, and Y. He. Does motivation in citizen science change with time and culture? In *CSCW Companion '14*, pages 229–232, New York, New York, USA, 2014. ACM Press.

[27] D. Rotman, J. Preece, J. Hammock, K. Procita, D. Hansen, C. Parr, D. Lewis, and D. Jacobs. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 217–226. ACM, 2012.

[28] R. Simpson, K. R. Page, and D. De Roure. Zooniverse: observing the world's largest citizen science platform. In *WWW '14: Proceeding of the 23rd international conference on World Wide Web*, pages 1049–1054, New York, New York, USA, Apr. 2014. ACM.

[29] J. Surowiecki and M. P. Silverman. The Wisdom of Crowds. *American Journal of Physics*, 75(2):190, 2007.

[30] Y. R. Tausczik, L. A. Dabbish, and R. E. Kraut. Building loyalty to online communities through bond and identity-based attachment to sub-groups. In *CSCW '14: Proceedings of the ACM 2014 conference on Computer Supported Cooperative Work*, pages 146–157, New York, New York, USA, 2014. ACM Press.

[31] H.-L. Yang and C.-Y. Lai. Motivations of Wikipedia content contributors. *Computers in Human Behavior*, 26(6):1377–1383, Nov. 2010.

[32] F. Zhang, N. J. Yuan, D. Lian, and X. Xie. Mining novelty-seeking trait across heterogeneous domains. In *WWW '14: Proceeding of the 23rd international conference on World Wide Web*, pages 373–384, New York, New York, USA, 2014. ACM Press.

[33] H. Zhu, R. E. Kraut, and A. Kittur. Organizing without formal organization: group identification, goal setting and social modeling in directing online production. In *CSCW '12: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, New York, New York, USA, 2012.