Linguistic Adoption in Online Citizen Science: A Structurational Perspective

Completed Research Paper

Corey Brian Jackson
University of California, Berkeley
Berkeley, CA
coreybjackson@berkeley.edu

Carsten Østerlund
Syracuse University School of Information Studies
Syracuse, NY
costerlu@syr.edu

Mahboobeh Harandi
Syracuse University School of Information Studies
Syracuse, NY
mharandi@syr.edu

Dhruv Kharwar
Syracuse University School of Information Studies
Syracuse, NY
djkharwa@syr.edu

Kevin Crowston
Syracuse University School of Information Studies
Syracuse, NY
crowston@syr.edu

Abstract

For peer-production projects to be successful, members must develop a specific and universal language that enables them to cooperate. Complicating the development of language in some projects is the lack of formalized structures (e.g., roles) that communicate to members the norms and practices around language. We address the question of how do role differences among participants interact with the adoption and dissemination of new terminologies in open peer production communities? Answering this question is crucial because we want communities to be productive even when self-managed, which requires understanding how shared language emerges. We examine this question using a structurational lens in the setting of a citizen science project. Exploring the use of words in the Gravity Spy citizen science project, we find that many words are reused and that most new words that are introduced are not picked up, showing a reproduction of structure. However, some novel words are used by others, showing an evolution of the structure. Participants with roles closer to the science are more likely to have their words reused, showing the mutually reinforcing nature of structures of signification, legitimation, and domination.

Keywords: Peer production, language use, citizen science, structuration, roles

Introduction

Peer production communities such as Wikipedia, free/libre open-source software (FLOSS) projects, and citizen science projects rely on the voluntary contributions of members of the public. In Wikipedia, volunteers write and edit articles; in Linux projects, both novice and expert programmers write and debug software code. For peer production projects to be successful, new members must join and learn how to collaborate towards a common goal. Newcomers must learn the norms of the group, in particular, its linguistic practices, to become productive members creating novel solutions to difficult problems. Language is an important medium for communication and even more so in online groups. Studies reveal that the level of enculturation into a group's language practices can predict an individual's success on the job (Goldberg, Srivastava, Manian, Monroe, and Potts 2015) and socialization into an organization (Goldberg & Srivastava 2017). Moreover, learning communal language practices supports coordination (Crémer et al. 2007), builds shared meaning and forestalls communication breakdowns (Bjørn & Ngwenyama 2009).

In traditional organizational settings, to evaluate a candidate's fit for the job, newcomers go through screening processes such as interviews and background checks. Moreover, once hired, an employee may undergo extensive training, sometimes lasting months, where, in addition to formal socialization, newcomers may observe the communication practices of more tenured employees. Together these socialization opportunities help newcomers assimilate but also gives them a foundation to produce new knowledge benefitting the organization.

Peer production faces, in contrast, a myriad of arrangements that make linguistic socialization difficult. First, many peer production communities have loose entrance requirements, meaning that participants join the community with linguistic differences. Participants will learn the language at various speeds depending on their prior experiences and commitment to the project. Moreover, new volunteers continuously join the project, further compounding the potential for linguistic differences in the community. Second, limited organizational control in peer production (Crowston 2011) often lead to uncertainty among newcomers as to what they need to know and the strategic direction of the organization. They may struggle to contribute to the development of new terminology and knowledge.

Despite these challenges of diffuse language competencies, limited organizational control, and lack of formal power structure, many peer-production communities successfully coordinate activities, their members communicate and they develop specialized terminology to address novel problems. Consider, for example, the many Wikipedia policies that guide editing and their associated specialized language, for instance, the commonly-used acronym NPOV for a "neutral point of view" (Kriplean et al. 2007). Volunteers developed these specialized terms and associated documents that help sustain the community and give participants a sense of direction (Arazy et al. 2015). However, learning the terms and the meaning behind them can be a barrier to entry for newcomers (e.g., Schneider et al. 2013). If a system's designers hold no formal power, how do newcomers know on which linguistic practices and specialized terms to focus their attention? Furthermore, if some peer-production communities explicitly require members to develop specialized terminology, how should these developments be organized and who is responsible?

To address these broader issues, the present study seeks to investigate the importance of power structures, communicated via roles (e.g., scientist, moderator, and volunteer), in shaping a community's language and in particular the development of specialized terminology. Specifically, in this article, we examine language adoption in a project called Gravity Spy, a citizen-science community in which volunteers classify glitches appearing in spectrogram images. In addition to classifying spectrograms, volunteers engage in conversations discussing the spectrogram, describing anomalies in the images, and debating and proposing labels for new kinds of glitches (Jackson et al. 2018). Gravity Spy places few restrictions on who can contribute to the discussion fora, only requiring volunteers to be logged in to post a comment. Beyond the classification task, the discussion boards and the norms associated with engaging in conversations are left to Gravity Spy's members. We also find that the loose governance structure in the discussion threads produces opaque power dynamics where volunteers may use personomies to labels glitches, i.e., develop terminology to describe the phenomenon. The lack of formalized governance structures leaves newcomers with little guidance about where to find and from whom to learn linguistic norms and practices.

In order to understand how innovative language practices and specialized terminology are produced and adopted in Gravity Spy, we study the project's discussion threads, focusing on how roles influence the adoption of words. Giddens' (1984) structuration theory is used as a lens to examine the production of language. Since many peer-production communities are run by the members themselves, there is a lack of formalized power structures to dictate norms and protocols. In contrast to traditional organizations with hierarchical governance structures, where employees know from whom to take cues, many peer-production communities have less visible, if any, power structures, which makes it difficult to know from whom to take direction and mimic behavior. We propose that the roles ascribed to individuals in the project offer a useful lens through which to determine how language is used, which words get adopted, and which words dissolve. The central question guiding this research is: *How do role differences among participants affect the adoption and dissemination of novel terminologies in open peer production communities?* Research from this viewpoint offers a way to understand how language in peer-production communities evolve, focusing on how structures of domination affect how words are used and adopted.

Structuration Theory

We use Giddens' (1984) structuration theory as a theoretical lens through which to examine the development of shared community language. A structurational perspective is suitable for this question as the theory describes a recursive relationship between structure and action, i.e., between a shared language and participants' contributions to the community that might be affected by or builds the language.

Structuration theory emerged as an alternative to functionalism, which holds that social structures are independent of people, existing in institutions, organizations, technologies, or other entities that can directly influence human behavior; and interpretivism, which holds that social structures exist only in the minds of people and hold no meaning outside of the social constructions that people create through their agency. Giddens argued that theories of structure or agency alone could not fully explain social action. Instead, we need a combination that considers the recursive relationship between structure and agency. On the one hand, actors operate within the context of rules produced by structures (traditions, institutions, moral codes, and established ways of doing things). But on the other hand, those structures are the result of previous actions and have no inherent existence or stability outside human action. They are socially constructed: only by humans acting in a compliant manner are these systems active and visible. Further, through the exercise of reflexivity, human agents can modify social structures by acting outside the constraints of the existing structure. In other words, Giddens argues that structures are neither dominated by institutions nor inherent to people; they are rooted in institutions and humans and occur in the actions people take, especially as they interact with one another.

Giddens names the interplay between the reification and transformation of structure structuration. Illustrated in Figure 1, structuration is a social process involving the reciprocal interaction of human actions and structures of organizations. Structuration occurs as actors invoke existing structures, producing and reproducing the structures and the associated social system, or act in ways to change the structures. The routine of everyday structuration constitutes the social order of a system. As a general theory of social organization, structuration theory helps explain the dynamic relationship between structures and action where the latter both reproduce existing structures but also produce new ones. In other words, members' actions can converge, reproducing stable structure or produce new practices pushing the limits of existing structures.

Structures can be analyzed into different categories, such as resources (command over people or material goods) and rules (recipes for action), which operate to provide a social system with power (structures of domination), norms/routines (structures of legitimation), and shared meaning (structures of signification). It should be noted that this division is an analytic convenience; in practice, these forms are integrated and mutually reinforcing. For example, structures of signification may be reinforced by structures of legitimation or domination that make certain interpretations more legitimate or required.

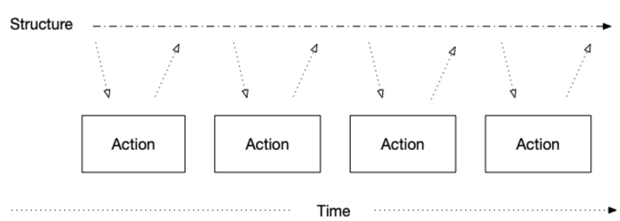

Figure 1. The interplay between structure and action over time described by Giddens' structuration theory.

Structuration and Language in Online Communities

In this paper, we examine shared language use as an indication of the creation and action of structures, specifically structures of signification. Communities in general and online communities in particular, invent specific modes of discourse that must be learned for participants to be able to contribute productively. Our perspective on linguistic structuration is illustrated in Figure 2. Community members read content posted by other volunteers from which they may learn the community language (structure) (the downward arrows in Figure 3). They may also interact with other material on the site (e.g., FAQ pages) from which they can learn the language used and what is expected of members.

When participants post content, they may reproduce existing language by posting content (action) whose composition is guided by what they have read, thus reproducing the existing language (the upward arrows in Figure 3). Such convergence in the language is vital for performance. For instance, Srivastava et al. (2018) found that employees who show a rapid convergence of their language with the language of the company were less likely to be fired and those whose language diverged after converging were more likely to quit, showing the importance of such reproduction of language.

However, participants may also create content that introduces their own ideas and approaches to communicating, perhaps different from past practices, thus adding their own voices to the community in the form of new terminology. Volunteers may be differently motivated because they bring varying competencies, intentions, and knowledge about the language and subject matter to a project.

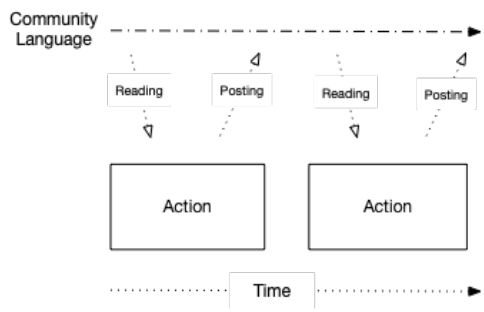

Figure 2. The interplay between structure and action in online communities.

These novel contributions of terminology may or may not influence future actions. It could be only that the contributions are too buried in the mass of content produced to be noticed. In many systems, most contributions are seen only by a few other community members due to the high volume of posts. Making contributions visible to others may require luck or a concerted effort by others. And even if it is seen, the contribution may not be emulated, again meaning that the structure does not change. However, contributions may occasionally be adopted by others, leading to a change to the structure and a further cycle of structure and action.

Jackson et al. (2018) describe the challenges associated with building a shared language in Gravity Spy, a peer-production project where volunteers can classify images (called glitches). The existing scheme has twenty labels; however, some images have a novel appearance, meaning the current scheme needs to be expanded to account for new glitch types. Volunteers are asked to innovate by creating new terminology. Volunteers do so by applying hashtags on a discussion thread unique to each image; creating innovative descriptive terms (e.g., *koilike*, the portmanteau of words koi and like to denote a morphology that resembles a morphology in the existing glitch in the schema called *koifish*), conventional terms (e.g., *1080line* to indicate a potentially class of glitches), and other words which are a part of standard lexicon. A challenge for expanding the schema is that volunteers enact personalized schemas, called personomies, to describe glitches, and these terms may conflict with the coinage of other volunteers. For example, one user may label a glitch *lighmodpaireddove*" while another user may tag it *paireddovelightmod*. A newcomer exposed to these two labels on the discussion thread may not know which label to use themselves, as there is no obvious way to facilitate label reconciliation.

One factor that can influence the uptake of an innovation is the status of the poster. Some online communities have distinguished roles for participants, e.g., a discussion group moderator, a committer in a FLOSS project, the various official roles in Wikipedia or being a member of a science team in a citizen science project. On Gravity Spy, participants' roles are listed underneath their avatar and username wherever it appears in the system, in the discussion threads in particular. Participants may have a role with some formal power (e.g., being able to edit a post or even ban a user), which create structures of domination (i.e., control over resources). A role may also be perceived as being more authoritative or responsible, which creates structures of legitimation. These structures can make contributions from participants in these roles more impactful and thus reinforce structures of signification.

We seek to understand how over time, members of peer-production communities produce and reproduce structures of signification in the form of specialized terminology and the role of power in this relationship. Examining this question in the case of online communities is interesting because the other structures of domination and legitimization that might shape the process are weak and so influence but do not determine the answer. Answering this question will provide insight into the way that these communities create or fail to create structures that enable productive collective action. Given the nature of the project as a science project, we hypothesize in particular that the adoption of words will depend in part on the legitimacy conveyed by the proximity of the poster to the science. However, in a voluntary community, structures are not determinative: others are free to ignore attempts to change the structure of language and participants in these roles. Powers of sanction exist but are a blunt instrument, wielded only in extreme cases. Thus, in addition to describing the production of language in the form of new terminology, we test the following hypothesis:

> **H1:** The proximity of a participant's role to the project's science increases the likelihood of a participant's word being adopted by other participants.

Our analysis extends the work presented in Jackson et al. (2018) in several important ways. First, we provide quantitative analyses of the terms. While they focused on comments and threads, we focus on the individual words that participants produce, a different level of analysis. Words were chosen because one goal of the project is for participants to create new words to label glitches, which Jackson et al. (2018) described as indicating structures of signification, i.e., what words are useful to refer to which classes of glitch. Second, we consider how roles (i.e., which incorporate structures of legitimation and domination) influence the patterns of word use and reuse in the project (i.e., structures of signification), something Jackson et al. (2018) did not consider.

Setting: Citizen Science

Our study is set in the context of a citizen-science project. Citizen-science projects that involve public participation in scientific research. People who contribute to citizen-science projects volunteer their time and efforts to help scientists collect, categorize, transcribe, or analyze scientific data (Bonney et al. 2009). Citizen-science projects can be performed entirely in-person, facilitated entirely over the Internet, or some combination of the two. In this paper, we focus on citizen-science projects facilitated entirely over the Internet (or online citizen science). Most online citizen-science projects ask volunteers to help analyze pre-existing data sets. In Galaxy Zoo, for instance, volunteers examine images of galaxies produced by the Hubble Telescope and ask volunteers to determine the morphological properties of the galaxy, choosing from a set of predefined choices, e.g., whether the galaxy shape is elliptical.

In addition to classifying data, some volunteers take to discussion boards to post comments to ask questions, share findings, or socialize with other volunteers. The commenting practices of volunteers have been described in several studies and suggest a variety of purposes for commenting, including communicating classification practices, sharing resources, welcoming newcomers to the community, alerting software developers to bugs in the code and discussing the underlying scientific focus of projects. The discussion has even led to novel scientific discoveries. For instance, the green peas galaxies and Hanny's Voorep were volunteer discoveries first shared on the discussion boards of Galaxy Zoo.

Shared language (i.e., structures of signification) are essential to online citizen-science projects in at least two ways. First, projects require interpretation of data, e.g., by classifying data into pre-given categories, which requires learning the meaning of the labels to be applied. Sometimes these are standard terms, like the animal species names in a project like Snapshot Serengeti, but in other cases, there is specialized terminology to learn, e.g., what counts as a transit in Planet Hunters. (Even in Snapshot, there are many unknown species to learn.) Second, participation in the discussion boards requires specific knowledge of the science behind the project and its specialized terminology.

Scholars note that reading and contributing to a citizen-science project's discussions help volunteers develop scientific literacy (Masters et al. 2016; Price 2011) as they observe conversations between knowledgeable participants, authoritative resources, descriptions of practice, and a community's social norms. Other studies note that learning about science motivate many volunteers to contribute to the projects and participate in the discussion boards (Raddick et al. 2009; Reed et al. 2013; Prestopnik and Crowston 2012; Nov et al. 2011).

A significant difference between citizen-science projects and many other peer-production communities is a definite status difference between participants, specifically between members of the science team and the volunteers. Through continued participation in the project, a new Wikipedia editor can aspire to higher roles, and a new FLOSS developer can aspire to be a core developer. A few citizen-science projects do provide additional opportunities for volunteers beyond just classifying (e.g., Crowston, Mitchell, and Østerlund 2019). However, a volunteer in a citizen-science project is quite unlikely to become a science team member, as to do so would require formal training in the underlying science in addition to acceptance by the team.

The status difference has implications for language use. In the sciences, professional scientists go through many years of training and socialization that shape their language and practices, enabling them to reproduce existing scientific structures. Citizen scientists typically lack this specialist background. As a result, they might not as readily reproduce existing structures through their actions, but instead, bring in new terminology or activities that could lead to changes in the structures. As noted above, citizen science projects often operate with comparatively weak forms of power, as members can join and leave as they desire. Science teams supporting citizen-science projects may strive to extend existing structures of signification (e.g., specific language) but they have few levers to control volunteers through structures of domination and legitimation.

Gravity Spy

Gravity Spy (Zevin et al. 2017) serves as the setting for this study, a citizen science project hosted on the Zooniverse platform (Simpson, Page, and De Roure 2014). Gravity Spy asks volunteers to help scientists search for gravitational waves by categorizing output from detectors used to search for gravitational waves. Due to the high sensitivity of the detectors, they record even the slightest noise from the environment or internal interference. Given the sensitivity of the detector, observation runs produce thousands of noise events (called “glitches”) which hinder the process of finding evidence of gravitational waves. Understanding the source of glitches and removing them from the data is vital to improving the sensitivity of the detectors.

The citizen scientist volunteers help scientists by classifying existing known glitches (useful to focus the search for the underlying cause) and develop potentially new classes out of new glitches. The glitch classification interface is shown in Figure 3. Volunteers scan the image and determine whether the noise pattern matches one of the twenty-two glitches in the current ontology, e.g., *whistle*, *koifish*, *tomte*. If the image does not match with a glitch in the list, volunteers have the option of selecting “None of the Above.” None of the Above spectrograms are important as they could potentially be examples of new classes of glitches. To improve the classification accuracy, each image is classified by multiple volunteers. Upon receiving sufficient classifications, the image is retired and a consensus glitch label applied to scientists.

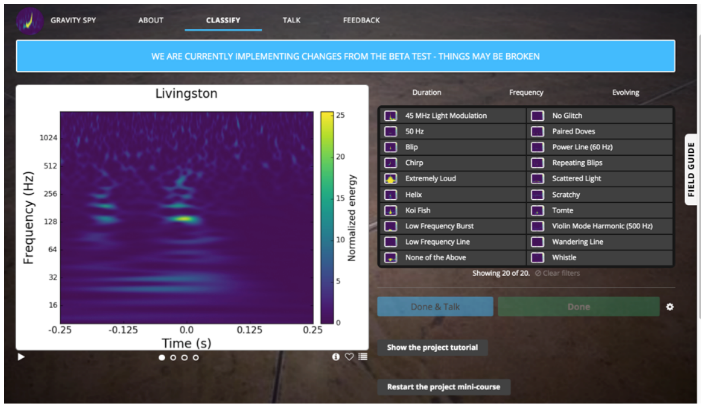

Figure 3. The classifications interface (left) where volunteers review spectrograms and select the appropriate glitch category from the options on the right.

While the classification of glitch data is the primary goal of the project, volunteers engage in conversations in Gravity Spy's discussion forums. While not a requirement, many volunteers post descriptions of their classification practices, share their terms for new glitches with other members, discuss the data, alert software developers of bugs in the system, and learn more about the underlying subject matter (Mugar et al. 2014; Jackson et al. 2015).

A specific use case for the discussion fora in Gravity Spy is to curate new glitches from the "None of the Above" category (Jackson et al. 2018). As a practice, volunteers often post a new label preceded by a hashtag to assign a name to the "None of the Above" glitch. Examples of newly proposed labels by volunteers include *1050harmonic*, *allfrequencies*, *elephantesque*, *electromagnetismo*, *rnahelix*, and *lighmodpaireddove*. Accompanying the label names are additional descriptions of the morphological characteristics of the glitch. For example, the comment "sub-category of blip with a 'bite' taken out of it" describes a new label proposed called *bitenblip*. More detailed comments may point to specific locations of the glitches in the image. For instance, in Figure 4, a more experienced volunteer responds to what appears to be a newcomer stating, "Maybe a few things going on here. Most prominent at t= 0.0 is what looks like a #aircompressor glitch. There is a good bit of the #70Hz line glitch and starting at around t= +0.125, maybe a #60HzPowerLine". The volunteer alerts future volunteers to a new glitch labeled *60HzPowerLine*, pointing out contextual information such as a similar glitch, i.e., *70Hz* which, if known can be used as a reference for where to look in the image and the location of the glitch in the image, i.e., t = +0.125.

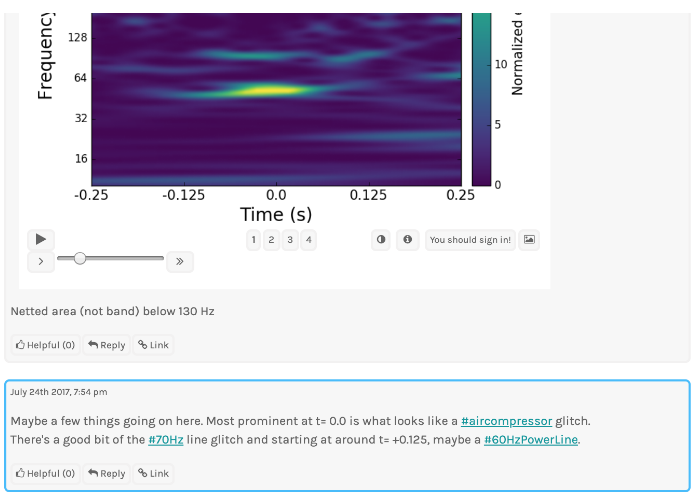

Figure 4. A screenshot of a discussion thread where one volunteer describes the location of potential new glitch and another volunteer follows-up with additional descriptions of the glitch and suggests the appropriate label might be #60HzPowerLine.

Methods

Below, we describe how we collected the data and our approach to data analysis.

Data Collection

To investigate how project language evolves, we gathered comments posted by volunteers in Gravity Spy from a February 2018 database dump. Given the bottom-up development of language in Gravity Spy, i.e., personomies and unique descriptions of glitches, our analysis focuses on the entire corpus of words. This means some words have been used by many volunteers of the project and some that have been used by only one or a handful of volunteers. In total, there were 68,969 records in the data dump. Each record in the database dump includes the body of the comment and attributes such as timestamp and username. The data includes comments posted by member of the science team, researchers, software developers, moderators, and volunteers.

Gravity Spy launched a test version in March 2016 and invited a group of volunteers to test the project while the software developers debugged the platform. The volunteers who joined during the testing phase contributed by regularly classifying data and posting comments to the system. The project's official launch took place on October 2016, after seven months of testing. We included comments from the testing phases of the project as setting the initial language visible to newcomers since the platform's discussion forums were not altered and individuals joining after the testing phase were able to view and reply to comments posted during the testing phase.

Data Analysis

To explore our research question, we consider how terminology in Gravity Spy is produced and reproduced by volunteers over time. In this paper, we start with the most straightforward analysis: reproduction and innovation in the specific words that volunteers produce. Since we are looking at new terminologies in Gravity Spy, we focused on unigrams and considered them equally important to reflect the community terminologies. For this analysis, we created tokenized words (unigrams) to represent language. To prepare the corpus of words for analysis, we performed standard text-mining procedures. Specifically, we removed stop words using the Mallet stop word list (McCallum 2002) and applied stemming to reduce each word to its root form. We retained numeric values since prior studies note their usefulness in communicating work practices (Mugar et al. 2014). While there were 68,969 posts in our initial dataset, 639 were removed because they contained only stop words. Additionally, the initial dataset contained 1,464 participants; the records of 16 volunteers were since their comments were only posts containing stop words.

In the results, we provide two types of analyses. We start with a descriptive summary of structure and how it changes over time. Second, we focus on structurational properties of language with an emphasis on how volunteers with different roles produce and alter language. Throughout the presentation of the results, statistics are conveyed using the following notations - mean (μ), standard deviation (σ), median (x). When models are presented, we use pseudo R^2 (Nagelkerke) to assess model performance.

A central variable in the study is the participant role, reflect both structures of legitimation (being scientists or being a recognized leading volunteer) and domination (i.e., enhanced abilities to manage data or the project). Some participants had multiple roles. For instance, one participant had expert, moderator, and researcher roles displayed beneath their username. We reduced multiple roles by choosing a single role based on its proximity to the actual science underlying the project. We assigned participants with more than one role to the following roles, from top to bottom:

- **Expert (N = 5)** - Participants who are members of the science team. “Experts can enter ‘gold mode’ to make authoritative gold standard classifications that will be used to validate data quality.”[1]
- **Researcher/collaborator (N = 21)** - Researchers are members of the science team, but without special access to the project. Collaborators have the ability to edit workflows and project content.
- **Moderator (N = 4)** - Moderators are active volunteers who are asked to help monitor the discussions and to answer questions and perform some additional functions on the discussion board, such as locking threads, moving threads and posts, editing thread titles and banning

[1] Roles are defined at https://help.zooniverse.org/getting-started/

volunteers, but have no power over the project itself. Moderators were first appointed by the Gravity Spy project team in Dec. 2016.

- **Ordinary volunteer (N = 1,418)** - Participants who are not assigned to one of the above roles are ordinary volunteers (henceforth called volunteers).

Results

We start by describing the corpus of words and the patterns of contribution of participants who post in the Gravity Spy Talk. Table 1 describes the dataset, which was produced over 1 year and 11 months (testing phase = 7 months, live phase = 1 year 4 months). The discussion boards are quite active, with more than 68,276 posts and 433,310 words contributed by volunteers. As is typical for online communities, participation is heavily skewed. 10,744 participants classified Gravity Spy data; however, only 13.3% of participants made a post during the data collection period (N = 1,432). Even among participants who did post, most contributed infrequently. Given the skewed distribution of participation behaviors in Gravity Spy, the median (represented as x) is a better representation of an "average" contributor. Table 1 shows that half of the volunteers posted three comments or fewer and their tenure as contributors to the discussion fora (the time from the first to last observed post) is only one day.

Table 1. Contribution statistics for all participants in Gravity Spy. Note: mean (μ), standard deviation (σ), and median (x)	
Observation Period	2016 March - 2018 Feb
Participants	1,432
Posts	68,276
Words (unique)	433,310 (17,720)
Participants (N = 1,432)	
	μ - σ - x
Days to first post	15.98 - 53.01 - 1
Posts	48 - 529.58 - 3
Tenure in forums	37.48 - 101.16 - 1

Table 1. Contribution statistics for volunteers in two projects. mean (μ), standard deviation (σ), median (x). Note: Participants include ordinary volunteers, experts, moderators, and collaborators.

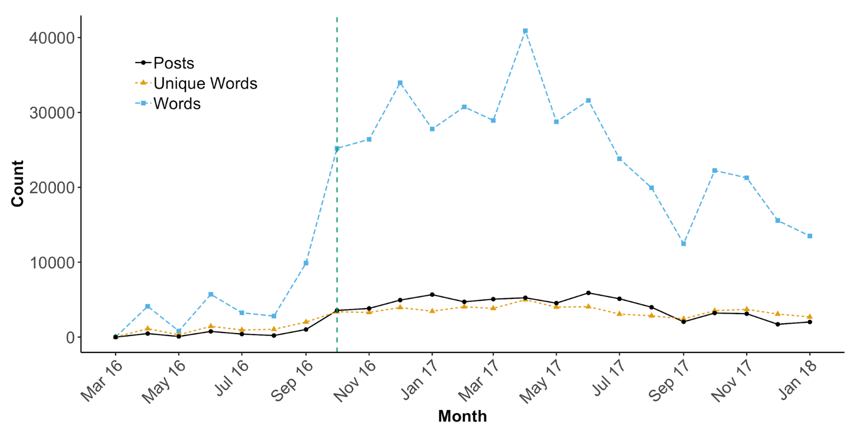

Figure 5. The quantity of posts, words, and unique words over time.

Figure 5 shows counts of posts, unique words, and total words by month. During the testing phase, there were 128 (8.9%) participants who contributed 3042 (4.4%) posts. The official launch on 12 October 2016 is shown by the grey vertical dashed line. The LIGO detector was actively observing from 29 Sep 2015 to 12 Jan 2016 (“O1”), from 30 Nov 2016 to 25 Aug 2017 (“O2”) and restarted observation on 1 Apr 2019 (“O3”). Between observing runs, the detectors were being worked on to increase their sensitivity. The start of the Gravity Spy project was chosen to precede the start of the second observing run; the drop off in participation corresponds to the end of the run in Aug 2017, though the Gravity Spy project continued to classify already collected data.

Re-producing existing structures

Participants reproduce existing structures by reusing the project’s existing words. As a proxy for the existing structure, we look at the proportion of words in a volunteer’s posts that had already been used by other volunteers, that is non-novel words in their posts. Figure 6 and 7 illustrates the quantities and proportions of the existing language reproduced over time for participants in different roles. The red vertical line indicates the beginning of the live phase. The lines are smoothed to visualize trends over time better.

During the testing phase, on average, there were 25 ($\sigma = 25$, $x = 12$) contributors each month. 24.5% of the words in each month were reproduced from the project’s existing language. In part, the high level of novelty of words is due simply to the project’s being in a startup phase (e.g., 100% of the words in the first post are considered novel). Usage increased after the launch: on average 131 ($\sigma = 66$, $x = 123$) participants contributed each month. Participants reproduced much more of the existing language: 47.7% of the words were reused on average. Still, the data show a high level of new words, reflecting in part the project’s goal of labeling novel classes of glitches, which requires new words.

Exploring these trends over time and by role suggests different relationships in reproducing structure. Figure 5 shows that since the beginning of the project, volunteers reproduce more of the project language than other groups. At the beginning of the project, the number of words rose steadily, reaching a pinnacle of 3,597 in April 2017 before dropping at the end of O2. The high volume reflects a large number of volunteers posting.

The moderators reproduced the second most words. Since moderators were only appointed in December 2016, their trend line is shorter than the other roles. The high volume is striking, given that there are only four moderators. For instance, in March 2017, three of four moderators reproduced 1,504 words, while 182 volunteers reproduced only 1,720. Experts infrequently contribute to the project discussions and thus the reproduction of its language. After the project's launch, two to three experts participated each month, reproducing 3,164 words, which was on average 58% of their words.

Collaborators also made infrequent contributions to the existing structure. Collaborators' interaction in the project declined dramatically after the project's launch, having an average of 2.26 ($x = 1$) posts per month. When collaborators posted, 32% of their words were reproduced from existing language; however, they did not post much with an average of only 27.66 ($\sigma = 27.8$, $x = 18$) words each month.

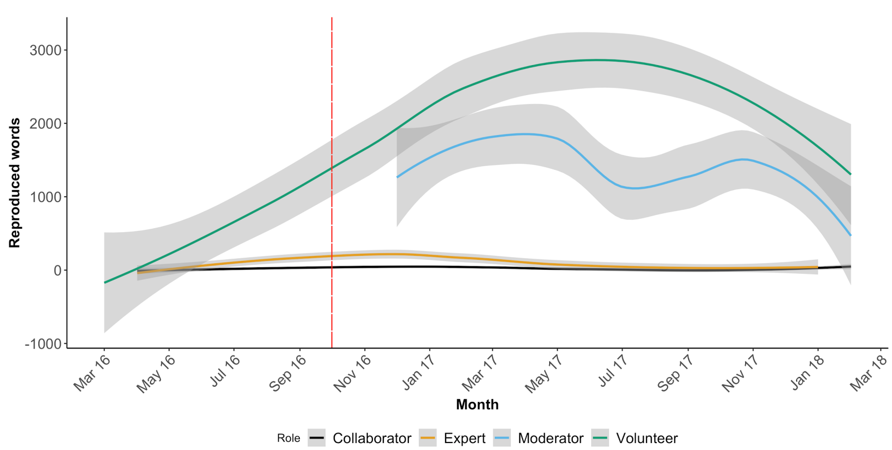

Figure 6. A smoothed line representing the count of words contributed by volunteers that were a part of the existing structure.

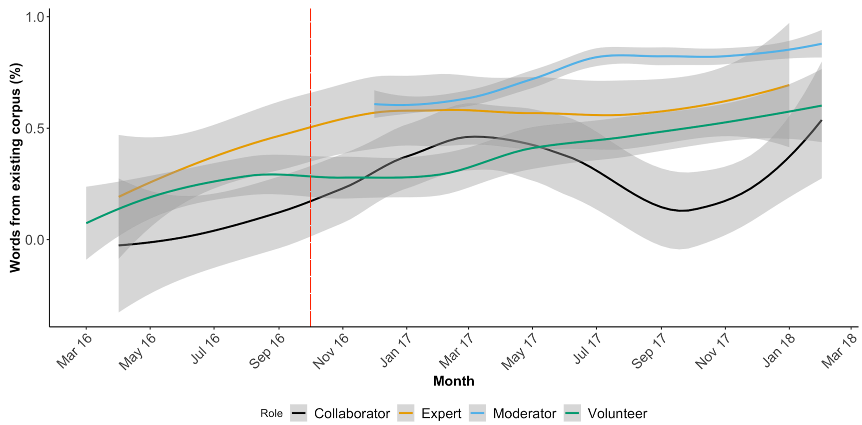

Figure 7. A smoothed line representing the proportion of words contributed by volunteers that were a part of the existing structure.

Figure 8 shows the mean proportion of words reproduced for each group, starting with their first month of contribution. The data are grouped by a participant's role in the project. The first moderator assumed their role in the second month as a contributor, so the first month of moderator status represented in the data was from the second month of participation. We also see that participants in some roles do not contribute in all months. For instance, collaborators participated in the discussion for only 12 months.

Most participants, regardless of role, start with between 21% and 27% of their language representing the existing language; as tenure increases, however, a more substantial proportion of words are reproduced. Compared to other groups, moderators' words have a higher proportion reproduced from existing words suggesting that as moderators post they reproduce existing language. Most interesting are new volunteers, who also produce their terminology early in the project.

Figure 8. The proportion of words reproduced since a participant joined the project grouped by the participants' role.

Production of new structures

Successfully producing new structures in the form of new terminology in the project can be challenging since the theorized supporting structures of legitimation and domination are less visible in Gravity Spy. Just using a new word is not sufficient to establish a new structure. This limitation is evidenced by examining which words are adopted in the project. While there were 17,713 unique words found in our data, 8,525 (48%) words were used just once in the project, that is, a participant may see some data, decide to place a new hashtag and never use the label again. We also find that more than half of the words were used only by the person who introduced the word: of the 17,713 unique words, 10,306 (58%) were used exclusively by one participant.

Introducing new structures

In this section, we analyze which participants influence the structure by introducing words and which participants support the permeation of words by being the next person to use the word. Figure 9 shows the mean number of words introduced in each month grouped by the participants' role. There were 14,000 (80%) words introduced during the live phase and 3,713 (20%) during the testing phase. 678 volunteers contributed new words to Gravity Spy. However, the contribution was again skewed. On average, volunteers introduce only a few new words; x = 3 ($\mu = 26.13$, $\sigma = 161.90$), while a few introduce a lot. Volunteers introduced 6,139 (83%), moderators 655 (9%), experts 550 (7.4%), and collaborators 65 (< 1%).

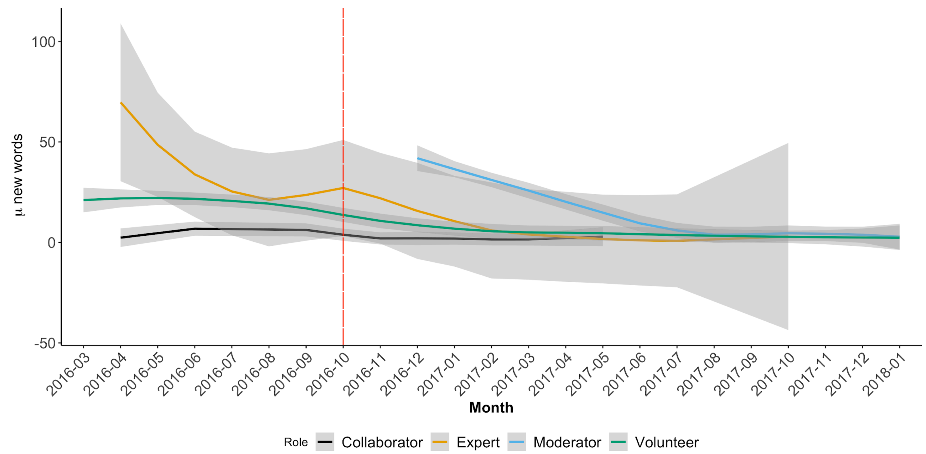

Figure 9. A line chart by participants' roles in the project showing the mean number of words that are new.

Perpetuating new structures

As noted above, introducing new words to the project does not guarantee that the words will be adopted. Adoption requires other participants to discover the word, understand, and appreciate its meaning, recognize cases where applying the word may be applicable, and choose to reuse the word. We computed the proportion of words that were adopted by volunteers each month based on their roles. Figure 10 shows for the subset of adopted words, the role of the individual who next uses the word, thus turning the word from a personomy to a folksonomy. We see that during the testing phase of the project, volunteers and experts played an important role in adopting words. However, some months after the project launch, the role of experts is diminished, and moderators and volunteers assume important positions adopting new words.

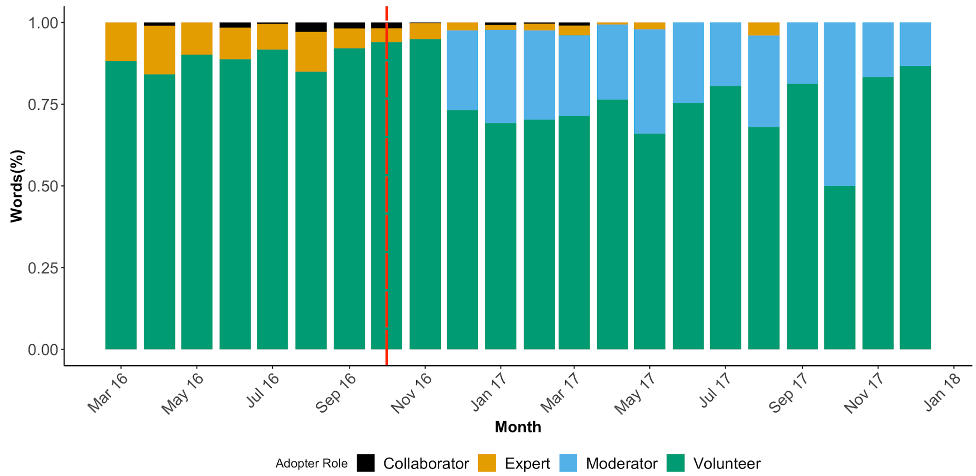

Figure 10. A stacked bar chart showing the role of the participants' who adopted words in the project over time.

Effects of role on structure changes

In our theorizing above, we noted that structures are mutually reinforcing, so we can look at how structures of legitimation or domination, even weak ones, support structures of signification. We do so by considering how roles figure into the adoption of words. We consider a word adopted if at least three other volunteers use it. In our dataset, of the 17,713 words introduced after the project launch date, 5,155 (29%) were adopted. Table 2 shows the results of a logistic regression model to predict word adoption. The model was trained on a 70% subset of the data and validated on the remaining 30% training. The model was a good fit: $\chi2$ (6) = 1978.2, $p < 0.001$ and showed a small (+1.1%) improvement over the no-information of 78.9%. The Nagelkerke (Cragg and Uhler) R^2 is 0.29, suggesting moderate performance.

Table 2. Logistic regression word *adoption*

	β	***SE***	***O.R.***	***p-value***
Intercept	-5.663	0.17	0.036	***
Innovator Use	0.02	0.007	1.02	**
Days Appearing	0.012	0.0003	1.01	***
Innovator's Tenure	-0.0007	0.0026	1.00	**
RoleCollaborator	0.66	0.41	1.93	
RoleExpert	0.49	0.17	1.63	**
RoleModerator	-0.17	0099	1.19	
Observations	12,400			
Log Likelihood	-989.12			
pseudo R^2	0.29			

The regression shows that the more occasions a person used a word prior to its possible adoption by another participant increased the likelihood of word adoption. On average, linguistic innovators used a word 1.89 (σ = 3.70) times before it was adopted. For each use by the innovator, the odds of the word being adopted by another volunteer increase by 1.02 (or 2%). Some words were used quite often prior to adoption. For instance, the word *lavalamp* was used by one volunteer on 119 occasions during an 8-day span before another participant used the word. Additionally, for each day the word had been used in the project, the odds of the word being adopted increased by 1.01 (or 1%). For instance, *thinblip* was used on 33 occasions over 51 days before it was adopted by another participant. Surprisingly, the number of days ago the participant joined the project (i.e., tenure) did not have a significant effect on adoption.

We turn next to the influence of role on the adoption of words. The regression shows that compared to a volunteer, if the linguistic innovator was an expert in the project, the log odds of the word being adopted increases by 1.63. We conducted follow-up tests to determine whether role pairs had different likelihoods of having their words adopted. The results of the chi-squared (χ^2) test showed significant differences among roles ($\chi^2(3) = 12.8$, $p < 0.05$. However, only experts and moderators were different, $\chi^2 = 3.4$, $p < 0.05$.

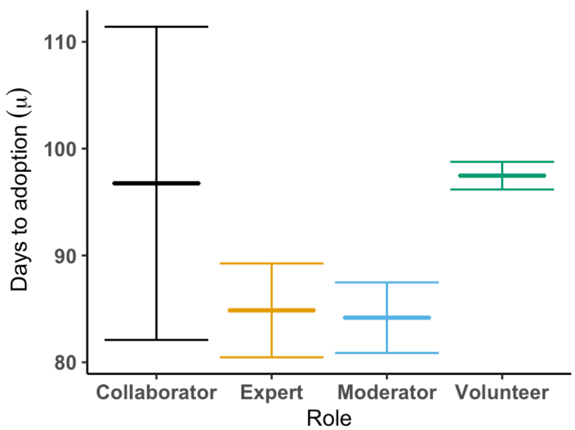

Figure 11. The mean time to word adoption by participants' role in the project

Finally, we sought to determine whether role affected the speed of word adoption. Figure 11 shows for each role the mean time of word adoption. On average, volunteers' words were adopted after 97 days ($\sigma = 101.3$, .70, $x = 66.33$), moderators, 84 days ($\sigma = 84.5$, $x = 57.37$), experts, 85 days ($\sigma = 103.1$, $x = 58.1$), and collaborators, 97 days ($\sigma = 118.2$, $x = 57.8$). We conducted an ANOVA to determine whether word adoption differed based on role. There was a significant difference of role on adoption time $F(3,7405) = 5.644$, $p < 0.001$. The Tukey follow-up test revealed only volunteers' adoption times were different from those of experts (-12.6, $p = 0.02$) and moderators (13.2, $p = 0.006$).

Discussion

The central question driving this research is to understand the development of new terminology among participants in an open peer production project with weak supporting structures for the adoption of a shared language. We explored this question through the lens of structuration and focused explicitly on the reproduction of existing language, the production of new structures over time, and how participant roles affect the adoption of words in the community.

When exploring the number of words, new words, and posts, volunteers continue to contribute in ways that both reproduce and innovate language. The results suggest continuous linguistic growth with new posts submitted regularly and new terms emerging to increase the project's linguistic base. However, people's ability to change structure is limited, as the adoption of new words can take time, and many new terms are never adopted. We find that participants with visible roles in the project predominantly facilitate the development of new terminology.

Giddens' structuration theory helps illuminate these processes. The evidence shows that the practices of participants with visible roles appear to influence the structuration process the most. The structurational perspective suggests that these roles add legitimacy to a participant's proposals for novel terminology. For instance, moderators' visibility and relative proximity to the science team appear to add legitimacy to their terminology proposals. Their visible roles allow them to significantly shape the project's shared language and meanings by making certain words and the interpretations they facilitate more legitimate or even required. For instance, some moderators have had their names for new glitch classes added to the pool of known glitches that all participants will see on the project interface. As their language practices get adopted among other members, the reuse reinforces structures of signification, thus influencing future classification practices.

The findings have implications for the new work arrangements emerging around open peer-production communities, which often lack mechanisms to support the establishment of shared meaning (signification) around the use of certain words or processes for volunteers themselves to introduce new terminology (legitimation). The findings are relevant across several tagging sites, e.g., Flickr, where we find users developing a new language to describe the phenomenon in digital media, e.g., images and videos. On BeerAdvocate, an online community where users evaluate and discuss beers, users develop new terms and conventions not conventionally used to describe beers, i.e., *gummy*, *corn-like*, or *overripe* (Danescu-Niculescu-Mizil et al. 2013). Our results suggest that project organizers pay attention to how they differentiate project roles and make them visible to the community. While the community has few enforcement mechanisms, the visibility of roles and their proximity to one another may help establish some structures of legitimation, signification, and domination that more effectively can shape the direction of the project.

The need to master this structuration process comes with some urgency. As open peer-production communities mature, so does their volume of new terminology. For instance, newcomers joining Gravity Spy today will face a significantly more complex language than a newcomer joining at the start of the project. There are a lot more discussion posts with a much newer terminologies to sift through. If volunteers do not know which words to apply to glitches, how to talk like an experienced contributor, they may cease contributing to the discussion boards or leave the project altogether. Mastering these structuration process may help address learning challenges associated with peer production projects.

Limitations

There are several limitations to the study. First, the dataset of single words may contain misspellings that could contribute to the larger linguistic corpus. Second, some words may be acronyms. For instance, *noa* is a common way to represent none of the above response. Third, moderators were selected partly based on their discussion activity, so the high volume of posts by moderators is partly cause as well as effect. Fourth, there are more to language structuration processes than the simple development of new terminology. Our focus on the development of specialized terminology offers one step towards a deeper understanding of the structuration process in open production communities.

Conclusion

Peer production communities rely on the public for participation. These communities often engage in the production of, not only new organizational forms but also new knowledge and with it new terminology. FLOSS teams develop code, Wikipedia has created a highly specialized terminology around editing, and citizen-science project members engages in the development of classification schemas and other forms of research. As these communities mature, the sheer volume of specialized terminology associated with each of these organizations has grown exponentially. A newcomer to Wikipedia today must learn a lot more terminology and specialized language than somebody joining around 2000.

Two factors complicate the development, sharing, and learning of such specialized terminology in peer-production communities. First, one finds a considerable span in language competencies. Second, these new organizational forms hold limited formal power to control and shape the practices of their participants and with it their production of new terminology central to the success of the projects. For instance, prior research by Jackson et al. (2018) described how volunteers struggled to introduce new glitch labels because the project lacked sufficient mechanisms to determine the most productive terminology and effectively promote their use across the community.

This paper expands on that finding by quantitatively assessing the evolution of new specialized terminology and analyzing how participants produce novel terminology about new glitch classes and reproduce existing language.

Acknowledgements

This work was supported in part by the U.S. National Science Foundation under grant INSPIRE 15-47880. We thank the many citizen scientists whose engagement have made Gravity Spy possible and our collaborators on the project. Our gratitude also goes to the review team whose comments have improved this paper.

References

Arazy, O., Ortega, F., Nov, O., Yeo, L., and Balila, A. 2015. "Functional Roles and Career Paths in Wikipedia," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 1092–1105.

Bjørn, P., and Ngwenyama, O. 2009. "Virtual team collaboration: building shared meaning, resolving breakdowns and creating translucence," *Information Systems Journal* (19:3), pp. 227–253.

Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T, Rosenberg, K.V., and Shirk, J. 2009. "Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy," *BioScience* (59:11), pp. 977–984.

Crémer, J, Garicano, L., and Prat, A. 2007. "Language and the Theory of the Firm," *The Quarterly Journal of Economics* (122:1), pp. 373–407.

Crowston, K. 2011. "Lessons from Volunteering and Free/Libre Open Source Software Development for the Future of Work," in *Proceedings of the IFIP Working Group 8.2 Conference*. Turku, Finland: Springer.

Crowston, K., Mitchell, E., and Østerlund, C. 2019. "Coordinating Advanced Crowd Work: Extending Citizen Science," *Citizen Science: Theory and Practice* (4:1).

Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., and Potts, C. 2013. "No country for old members: User lifecycle and linguistic change in online communities," in *Proceedings of the 22nd international conference on World Wide Web*, Rio de Janeiro, Brazil. pp. 307–317.

Giddens, A. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley: University of California.

Goldberg, A., Srivastava, S., Manian, V. G., Monroe, W., and Potts, C. 2015. "Fitting In or Standing Out? The Tradeoffs of Structural and Cultural Embeddedness," *American Sociological Review*, (*81*:6), pp. 1120–1222.

Jackson, C, Crowston, K., Østerlund,C. and Harandi,M. 2018. "Folksonomies to Support Coordination and Coordination of Folksonomies." *Computer Supported Cooperative Work (CSCW),* (27:3), pp. 647–678.

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

Kriplean, T., Beschastnikh, I., McDonald, D. W., and Golder, S. A. 2007. "Community, Consensus, Coercion, Control: Cs* w or How Policy Mediates Mass Participation." in *Proceedings of the 2007 International ACM Conference on Supporting Group Work*. pp. 167–176.

Masters, K., Oh, E.Y., Cox, J., Simmons, B., Lintott, C., Graham, G., Greenhill, A. and Holmes, K. 2016. "Science Learning via Participation in Online Citizen Science." *ArXiv Preprint ArXiv:1601.05973*.

McCallum, A.K. 2002. "Mallet: A Machine Learning for Language Toolkit."

Mugar, G., Østerlund, C., Hassman, K.D., Crowston, K. and Jackson, C.B. 2014. "Planet Hunters and Seafloor Explorers: Legitimate Peripheral Participation through Practice Proxies in Online Citizen Science," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. pp. 109–119.

Nov, O., Arazy, O. and Anderson, D. 2011. "Dusting for Science: Motivation and Participation of Digital Citizen Science Volunteers," in *Proceedings of the 2011 IConference*. pp. 68–74.

Prestopnik, N.R. and Crowston, K.. 2012. "Citizen Science System Assemblages: Understanding the Technologies That Support Crowdsourced Science," in *Proceedings of the 2012 IConference*. pp. 168–176.

Price, C. A.. 2011. *Scientific Literacy of Adult Participants in an Online Citizen Science Project*. Tufts University.

Raddick, M.J., Bracey, G., Gay, P.L., Lintott, C.J., Murray, P., Schawinski, K., Szalay, A.S. and Vandenberg, J. 2009. “Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers,” *ArXiv Preprint ArXiv:0909.2925*.

Reed, J., Raddick, M.J., Lardner, A. and Carney, K. 2013. “An Exploratory Factor Analysis of Motivations for Participating in Zooniverse, a Collection of Virtual Citizen Science Projects,” in *Proceedings of the 46th Hawaii International Conference on System Sciences*. pp. 610–619.

Schneider, J., Samp, K., Passant, A. and Decker, S. 2013. “Arguments About Deletion: How Experience Improves the Acceptability of Arguments in Ad-Hoc Online Task Groups,” in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW ’13. New York, NY, USA. pp. 1069–1080.

Simpson, R., Page, K.R and Roure, D. 2014. “Zooniverse: Observing the World’s Largest Citizen Science Platform,” in *Proceedings of the 23rd International Conference on World Wide Web*, pp. 1049–1054.

Srivastava, S. B., and Goldberg, A. 2017. “Language as a Window into Culture,” *California Management Review*, (60:1), pp. 56-69.

Srivastava, S.B., Goldberg, A. Manian, V. G. and Christopher Potts. 2018. “Enculturation Trajectories: Language, Cultural Adaptation, and Individual Outcomes in Organizations,” *Management Science* (64:3), pp. 1348–64.

Zevin, M., Coughlin, S., Bahaadini, S. Besler, E., Rohani,N., Allen,S., Cabero,M. et al. 2017. “Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science,” *Classical and Quantum Gravity* (34:6).