Shifting Forms of Presence: Volunteer Learning in Online Citizen Science

COREY JACKSON, University of California, Berkeley
CARSTEN ØSTERLUND, Syracuse University
MAHBOOBEH HARANDI, Syracuse University
KEVIN CROWSTON, Syracuse University
LAURA TROUILLE, Zooniverse

Open collaboration platforms involve people in many tasks, from editing articles to analyzing datasets. To facilitate mastery of these practices, communities offer a number of learning resources, ranging from project-defined FAQs to individually-oriented search tools and communal discussion boards. However, it is not clear which project resources best support participant learning, overall and at different stages of engagement with the project. We draw on Sørensen’s framework of forms of presence to distinguish three forms of engagement with learning resources: authoritative, agent-centered and communal. We analyzed trace data from the Gravity Spy citizen-science project using a mixed-effects logistic regression with volunteer performance as an outcome variable. The findings suggest that engagement with authoritative resources (e.g., those constructed by project organizers) facilitates performance initially. However, as tasks become more difficult, volunteers seek and benefit from engagement with their own agent-centered resources and community generated resources. These findings suggest a broader scope for the design of learning resources for online communities.

CCS Concepts: • Human-centered computing → Empirical studies in collaborative and social computing.

Additional Key Words and Phrases: citizen science, user behavior, learning

ACM Reference Format:

1 INTRODUCTION

Learning by doing characterizes many online production communities, such as Wikipedia, open-source software, or citizen science. Even so, these communities typically provide at least some resources to train and socialize new community members. Wikipedia has, for example, a set of pages for new editors that introduce policies and conventions governing participation on the site and describe how to style Wikipedia articles. Other pages provide more experienced members with best practices, e.g., on how to interact with newcomers [8]. Q&A communities describe how to contribute. For instance, StackOverflow provides new users with a two-minute tutorial covering how best to formulate questions and the benefits of applying tags. Communities might advertise...
best practices in frequently asked questions (FAQs) or about pages. In an online citizen-science project, Mugar et al. [19] found the comments left by users in the Planet Hunters project served as valuable learning resources for newcomers, as they pointed to specific features of work practice that were lacking in tutorials.

FAQ, how-to pages, and comments may be valuable resources for learning and socialization, but we know little about the process by which members of the community make use of them over the course of their tenure on a project. Identification of the assemblage of resources (or structured patterns of use) adopted by users helps us see how they make sense of their environment and learn. Knowledge about which resources are useful as participants learn to contribute could help those who manage online communities know which artifacts to provide or to suggest to users. A complication is that different resources may be useful at different points in the learning process. For example, a tutorial could be a valuable learning resource for newcomers during the beginning stages of participation but lose its significance over time.

In this paper, we draw on trace data from an online citizen-science project to examine the resources that support participant learning to perform a data analysis task. Citizen-science projects engage members of the public in scientific research. While there are several models of citizen science, the project we investigate here involves volunteers in large-scale scientific data analysis. Such citizen-science projects rely on an online worldwide collaboration platform to support the involvement of scientists and the public. The scientists share their research projects with the public who are interested in science. However, as the volunteers may not have relevant background knowledge, scientists typically also provide multiple learning resources to educate the volunteers. As a result, citizen-science projects can be as informal learning opportunities for contributors [2].

Specifically, we present an analysis of resources used by volunteers in Gravity Spy, a citizen-science project hosted on the Zooniverse platform. A distinctive feature of the system is that it collects data on the accuracy of the volunteers on the citizen science task, providing an opportunity to assess how well different volunteers have learned the task. Using this data, we can compare how the use of different kinds of learning resources (e.g., tutorials, FAQs, forum posts) at different stages of participation affect performance. Specifically, we address the question: Which project resources seem to best support user learning at different stages of engagement with the project?

2 LEARNING RESOURCES AND FORMS OF PRESENCE

With the proliferation of online citizen science, blogging, video sharing, and MOOCs, we find a fast-growing literature on learning in such crowdsourced and peer-based platforms. Many studies have a strong educational focus [6, 10, 11, 15, 17, 22], but we also find a growing body of work focusing on learning in crowdsourced and other online peer-production settings, e.g., Amazon Mechanical Turk, Wikipedia, Zooniverse or YouTube. A large part of the peer-production based work builds on a practice-based understanding of learning, which does not pin knowledge to the heads of individuals but situates it in a social and material context; whether conceived as activity systems [7], communities of practice [16], sociocultural [25], or socio-material formations [26]. Learning emerges as participants gradually expand their engagement with the resources on the site. These resources can take many forms spanning from explicit educational materials to traces of prior activities on the site, and engagement with other participants and experts. Engagement with peer production project resources support participants’ learning and thus their ability to do the required work.

It is less clear, however, when in the learning process engagement with different kinds of resources is particularly helpful. We know from the literature that people’s practices and learning changes over time. Several influential concepts pertain to this process. The notion of legitimate peripheral
Shifting Forms of Presence: Volunteer Learning in Online Citizen Science

participation highlights the importance of learners gradually gaining access to more and more communal practices [16]. The zone of proximal development concept emphasizes the need to adjust learning opportunities to the trajectory of specific learners [7]. Bringing these concerns together, the concept of scaffolding considers how one best sequence participants access to these various learning resources, whether new activities, materials, fellow participants, or experts [14]. In a study of Wikipedia, for instance, [3] shows how novices often start by simply reading other’s articles before they start making their initial contributions. Gradually they gain access to more involved tasks. Likewise, Preece and Shneiderman [24] suggests that participants in peer-production sites move from readers to leaders.

To characterize different kinds of learning resources that might be appropriate at different stages, we turn to work by Sørensen [26]. She distinguishes three forms of learning each associated with a different form of situated presence and engagement: (1) authority-subject, (2) agent-centered, and (3) communal presence. The question becomes when in a learning process each of these forms of presence and their associated types of resource engagement become more or less important. We will address each type in turn.

2.1 Authority-subject form of presence

First, when engaged in an authority-subject form of presence, the learning resources, expertise, and learners are divided into clear sub-regions associated with clusters of homogeneous activities and artifacts. The classic classroom serves as an iconic example with its division into two sub-regions: the front of the classroom, inhabited by the teacher and the blackboard, and the rest of the room, in which students face the teacher and the blackboard. In this format, teachers serve as authorities over the students. Learning happens when true and previously tested knowledge is transmitted from the teacher’s sub-region to the students, who will then imitate the teacher’s practices to reinforce the transmission. Authoritative resources such as textbooks further reinforce the clear distinction between authoritative knowledge and its subjects. In summary, the authority and the subject each occupying their specific positions in an infrastructure. The subject learns by imitating exemplary and prototypical knowledge displayed in the authority’s region. These forms of authority knowledge will often be chunked up and presented to the learner in carefully organized sequences.

In the crowdsourcing learning literature, we find several studies centered around clearly demarcated regions promoting authoritative resources to novice subjects. For instance, Mitra and Gilbert [18], Willett et al. [30] applied examples of expert solution of micro-tasks to train newcomers and calibrate their performances. Several scholars have found that feedback or examples of expert-classified data can serve as authoritative resources. Oleson et al. [21] used data as a form of training to transmit authoritative practices to newcomers. Likewise, Bateman et al. [1], Spohrer and Soloway [27], Walker and Engel [28] highlight the importance of expert feedback. In the context of citizen science specifically, researchers have found that expert-supplied resources (e.g., FAQ, field guides) are valuable promoters of learning [9, 13, 23].

There is evidence that user-generated resources can also support learning in peer-production settings [4, 32]. Doroudi et al. [4] find that participants learn by not only performing gold standard tasks and reviewing expert example solutions but also by validating solutions created by other workers. However, reviewing peer-work solutions was more effective for more advanced learners who engage these resources through their collaboration with other community members, rather than simply taking them as given.

Overall, it seems that authority-subject forms of presence are effective learning resources early in the tenure of a peer-production participant. Only later may participants gain the full benefits of other forms of presence and resource engagement. In other words, we hypothesize that newcomers will find authority-subject resources more useful compared to experienced participants and that
these authoritative resources will lead to better work performances early in participant’s tenure with a project.

- **H1a:** Participants early in their engagement with the project will use more authority-subject resources than agent-centered or communal resources
- **H1b:** Participants who use authority-subject resources will improve their performance on the task early in their engagement with the project

### 2.2 Agent-centered form of presence

Second, agent-centered presence and learning is associated with fluid relations. Individual agency directs participants’ activities, as one can observe if you let a group of 4th graders loose on Minecraft. They will move through the virtual environment trying one thing and then another, making connections and bringing disparate components together. This playful exploration is not guided by an outside authority or the collective, but rather by the practice. As such, it moves beyond the stable relations found in authority-subject and communal forms of presence and can be hard to scaffold. It is an aggregate process that builds over time where one or more participants can gradually build knowledge by adding to the infrastructure. The fluid learning typically associated with agent-centered presence involves ongoing mutations of the knowledge that the participants master and the spaces and times in which they do so. It is a process where the last step influences the next and thus gradually mutates over time as the participants extend their socio-material relations, and new spaces emerge.

Østerlund et al. [23] find similar learning processes among citizen scientists, who extend their exploration beyond one specific online citizen project to include many other online and offline resources. Sometimes these explorations go beyond the stated goal of a project. For instance, in the citizen-science project named Galaxy Zoo, one participant noticed images of galaxies that looked like green peas. Volunteers’ search for similar images eventually leads to the publication of an article describing this previously unknown astronomical phenomenon.

We hypothesize that agent-centered forms of resources become prevalent once a volunteer has gotten the hang of things and knows how to navigate and contribute to a project. In other words, we expect that more advanced participants will engage agent-centered resources and these will have a bigger impact on their performance in the projects compared to authority-subject resources more prevalent at earlier stages of their engagement.

- **H2a:** Participants later in their engagement with the project will use more agent-centered resources than authority-subject resources
- **H2b:** Participants who use agent-centered resources will improve their performance later in their engagement with the project

### 2.3 Communal form of presence

Finally, communal presence forms around a central collective activity, object or event. Learning associated with communal presence takes place as participants join the community and strengthen their relations to the center. Knowledge becomes validated through joint engagement in the communal practice. A prototypical example is a seminar, in which students learn through group discussion. Socially, they move toward the center of the community as they become sustained participants, increasingly fluent in the tasks, vocabulary and organizational principles of the community.

The crowdsourcing literature focusing on communal forms of presence often distinguishes collaborative activities, defined as developing relationships, working together and setting goals, versus contributions to the community, through the rating, tagging, reviewing, posting and uploading of
content [24]. Learning is enhanced in particular by the later, contributions in a communal setting. Weir et al. [29] coin the concept of learner-sourcing to describe the benefits learners gain from showing how they solved a problem, i.e., “show their work.” The process of explaining their actions help to solidify their understanding. At the same time, their work can benefit future learners by leaving descriptions of practices that can be observed in the discussion threads Mugar et al. [20]. To newcomers these traces act as proxies for expert practice, that is, they make visible the socially-salient aspects of people’s unfolding work practices without requiring the practices themselves to be shared. They help learners get a sense of expert work in situations where they may not have direct access to view other’s unfolding activities. Contributions can thus serve as user-generated learning resources, accessed as part of participation in the community. Thus, we hypothesize that:

- H3a: Volunteers later in their engagement with the project will use more communal resources than with authority-subject resources
- H3b: Volunteers who use communal resources will improve their performance later in their engagement with the project

At this point, we are left wondering about the relative importance of agent-centered and communal forms of presence and resource engagement. Neither Sørensen [26] or the peer-production literature offer any clear hints to whether one dominates the other or if they both increase their importance in tandem as volunteers mature. Given that many peer-production projects rely on individuals working alone in front of their computers it might seem that there are limited opportunities to interact with virtual peers. But many open-production platforms have fora in which participants can interact and rally around central collective activity, object or event. At the same time, they are free to roam and followed their own agent-centered activities. Rather than making specific hypotheses, we will empirically explore the relationship between these forms of presence.

3 PLATFORM: GRAVITY SPY

Our study is set in the context of a citizen-science project. In October 2016, Zooniverse, the LIGO Scientific Collaboration (LSC), and other researchers launched Gravity Spy [31], a citizen-science project to improve the interferometers used to search for gravitational waves. A challenge for LIGO scientists is the high sensitivity of gravity detectors, which is needed to search for gravitational waves, but which results also in recording a large quantity of noise events (referred to as glitches). The glitches obfuscate or even masquerade as gravitational-wave signals, reducing the efficacy of the search. Currently, there are more than 20 known classes of glitch with different causes, with the possibility of identifying more classes. Gravity Spy recruits volunteers to classify glitches, which helps to focus the search for their source. Figure 1 shows the classification interface.

Scaffolded Participation. A novel feature of Gravity Spy is an implementation of a training regime, in which new volunteers are gradually introduced to the glitch classes: first two, then five, and only after practice, all of the classes [31]. As a result, participation in Gravity Spy is scaffolded: volunteers start in Workflow 1 and based on their performance can advance to upper levels. Each workflow and the added glitch classes are shown in Table 1.

Importantly for our study, as volunteers classify the glitches, they are periodically given gold-standard data to classify (i.e., data with a known classification) in order to assess their accuracy at the classification task and readiness for promotion to a higher level. To determine when to promote a volunteer to a higher level and also to make a decision if an image (test data) should go to a higher level or be retired, the system applies a crowdsourcing classifier. As soon as a volunteer joins the project and starts the classification, the system makes a confusion matrix for them. The reliability model is built based on the probability of providing a true
**Frequently Asked Questions**

**Gravitational Waves FAQ**

**What are gravitational waves?**

Gravitational waves were predicted by Albert Einstein’s theory of general relativity in 1916. He showed that certain systems distorting spacetime as they move, sending out ripples in the fabric of spacetime itself. An (un)imagining your finger moving through water, if your finger is still, no waves will be created on the water if you start to move your finger through the bowl of water, ripples will propagate outward. Gravitational waves have the effect of stretching and compressing space and time.

**What is LIGO?**

LIGO stands for the Laser Interferometer Gravitational-Wave Observatory, the largest experiment in the universe. It is currently in operation. The LIGO Scientific Collaboration (LSC) consists of scientists from 8 institutions, all working together to develop the field of gravitational-wave astronomy. LIGO operates observatories in the United States, LIGO Livingston in Louisiana and LIGO-Hanford in Washington with the Virgo observatory in Italy.

**How do you detect gravitational waves?**

In essence, the instruments LIGO uses to detect gravitational waves are called Michelson interferometers.
Learning Resources on Gravity Spy. The Gravity Spy project site contains various pages through which volunteers can learn about the project from the organizers organize their work and interact and socialize (i.e., authority-subject, agent-centred and communal forms of presence). As examples of the first, the FAQ page (Figure 1) gives volunteers background information about the project, such as descriptions of gravitational waves, the LIGO collaboration and the science behind gravitational wave detection. The field guide shows each glitch category and describes the morphological characteristics of a glitch class, pointing to where the noise frequencies are most likely to occur. For instance, the description below the glitch selection describes blips as “short glitches with a symmetric ‘teardrop’ shape”. The field guide provides prototypes and a longer description,

“Blips are short glitches that usually appear in LIGO’s gravitational-wave channel with a duration of around 40 ms, at frequencies between 30 and 500 Hz, and with a symmetric ‘teardrop’ shape in time-frequency. Generally, they have more normalized energy at the lower frequency end of the teardrop, and their energy is overall modest compared with other glitch classes.”

The description and accompanying prototypes were selected because they represent a typical glitch of the class. However, some classes have more variability, meaning that glitches will not necessarily look exactly like the chosen prototype.

As an example of the second, the site supports collections for volunteers to keep track of images they find interesting. Collections are also a way for volunteers to organize their independent research activities. As an example, a user created a collection titled “Paired doves timing with alternate morphology” to investigate glitches that share the same 0.4 Hz timing as the paireddoves glitch, but with a different morphology and with a weak amplitude. The collection contains more than ten comments by other volunteers who are also interested in adding to the collection.

As an example of the third, the project hosts five types of discussion boards: science, notes, help, discussion and technical. Science boards contain conversations about the science behind gravitational wave research and related scientific fields. Figure 1 gives an example, a discussion of the phenomena of low-frequency bursts observed in one of the interferometers (i.e., Livingston). Notes are conversations about a unique image. When debate exists about what features of an image causes it to be classified as one glitch category versus the other, volunteers sometimes leave comments about their reasoning. As an example, in debating whether a glitch was a wandering line or a 1080 line class of glitch, one participant stated, “Most of the subjects in the collection have a hint of wandering Line mostly above the 1080line. They’re quite faint, but if you look closely, you may see the wavy, wandering pattern...”. Help and technical boards are for general questions related to contributing or for noting bugs in the interface. The discussion board is a general board for conversations about any subject.

4 METHODS

Our study is a mixed-method study involving both quantitative and qualitative data collection and analysis. The quantitative data for this research were collected in January 2017 and cover activities of Gravity Spy volunteers span one year and nine months. The qualitative data consist of interviews with power users.

4.1 Quantitative Data Collection

Our data originated from two sources. First, a system log containing the event history of all visitors to Gravity Spy. The earliest record in the system log was March 2016, while the most recent record was dated December 2017 (1 year and nine months). The system log contains records of volunteers...
interactions as they navigate the website. The system records events such as clicking the frequently asked questions (FAQs), opening the project field guide or accessing account information. In addition to the interaction, a timestamp is recorded with the exact date and time an event occurred. We also obtained a database dump of the classifications volunteers executed during the same time frame. The classification dump contained every classification a volunteer executed while logged in to their account. Each record in the data dump contained a unique identifier for the subject, whether the subject was gold, the volunteer’s answer and a timestamp indicating when the classification was posted to the system. To make the data ready for analysis, we conducted several data processing steps, which we describe below.

**Step 1 Merging data.** To construct a history of volunteers’ interactions in Gravity Spy, we combined the system logs and classification data dump. We simply added the classifications to the system logs and ordered the data by the timestamp.

**Step 2. Reconciling paths.** The data about volunteer actions on Gravity Spy requires extensive pre-processing. The system records the provenance of the interaction as well as the interaction itself. Specifically the system records different paths to a particular resource, depending on where the user is on the site. For example, the record when a volunteer is classifying and clicks view discussion will be represented as `classify_view_discussion` but when a volunteer is on the project landing homepage and clicks view discussion it will recorded as `about_view_discussion`. As a result, the data include different representations of what we treat a single kind of event (view discussion). Additionally, interaction-specific data may be included in the record. For instance, when searching, the text of the search is included in the database record, e.g., `blip_search`. Accounting for unique naming and interaction-specific data, we observed more than 1000 unique types of events; however, our reconciliation procedures resulted in 51 unique events.

**Step 3 Grouping data.** After reconciling events, two Ph.D. students collaboratively coded them to group them into our theoretical categories. This group was important both theoretically and also for analysis, as different events occur with different frequencies, making it difficult to see patterns in the data at the most detailed level. First, we read through each resource to determine whether it was a learning resource or an interaction. For instance, while the interaction logs recorded when a volunteer posts a comment, we did not consider posting to be accessing a learning resource. In total, we identified 43 learning resources Next, we examined each learning resource to determine the form of presence. Our rule for grouping data was simple. If project organizers constructed the page, we considered it authority-subject (N = 3), if the volunteer constructed the resource, we considered it agent-centered (N = 7), and if other volunteers constructed the resource, we considered it communal (N = 33). An list of the forms of presence with examples are included in Table 2.

<table>
<thead>
<tr>
<th>Form</th>
<th># Interactions - Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authority-Subject</td>
<td>3 - field-guide, tutorial, image metadata</td>
</tr>
<tr>
<td>Agent-Centered</td>
<td>7 - user-collections, user-favorites, search</td>
</tr>
<tr>
<td>Communal</td>
<td>33 - talk-board, member-profile, post comment</td>
</tr>
</tbody>
</table>

Table 2. The forms of presence and examples based on Sørensen [26].

4.2 Quantitative Data Analysis

The analysis is grouped into several sections. First, as background to our study of improvement in performance, we report on the overall performance on the glitch categorization task in the project, specifically performance overall and on different classes of glitch.
Second, we describe resource usage in the project for each form of presence described in the previous section. To be able to assess relative usage of different kinds of resources requires clustering the interactions into larger units. We aggregated the data into sessions. We define a session as a group of consecutive activities separated by no more than 30 minutes. The intuition behind the definition of a session is that volunteers often log in to the system, contribute for some time, and then take a break, e.g., until the next day. As volunteers do not always log out of the system when they are done classifying, we determine the end of a session based on timing. A gap of more than thirty minutes indicates the start of a new classification session. By comparing use of resources in initial and subsequent sessions, we test our hypotheses about how volunteers’ forms of presence change as they advance in the project.

Third, we examine the impact of resources of different types on the volunteers’ performance. We take advantage of a set of what are essentially on-going micro-experiments. While classifying, if a volunteer is shown a gold standard image and misclassifies the glitch, the system notifies the volunteer that their answer was incorrect, displaying a message reading “You responded Whistle. When our experts classified this image, they labeled it as a Blip. Some of the glitch classes can look quite similar, so please keep trying your best. Check out the tutorial and the field guide for more guidance.” (The specific classes mentioned change depending on the volunteer’s answer and the glitch.) We check which types of learning resource are used after the feedback until the next classification of the same type of glitch (e.g., if the volunteer consults the field guide as suggested). The outcome variable is whether the volunteer makes the next classification correctly (i.e., if they improve in accuracy). The specific test is a mixed-effects logistic regression model with participants included as a random effect to control for the non-independence of data at the user level since participants classify multiple images. Our model included main effects variables for the three forms of presence as binary variables indicating whether a volunteer used a resource of that type between making the two classifications.

4.3 Qualitative Data
To augment the quantitative analysis, we conducted interviews with four “superusers” of the Gravity Spy project, who we refer to by the pseudonyms Britney, Mike, Jason and Audrey. Interviews were conducted during two periods: the beta stage of the Gravity Spy system and in the month following the project’s live launch. Interviewees were selected based on their status as having executed a large number of classifications in the project. The purpose of the interviews was to understand how volunteers use resources to support their participation and how their use of resources changes over time. During the interviews, each interviewee mentioned aspects of their participation that helped them learn how to identify and curate glitches.

The interviews were semi-structured and lasted approximately one hour. All interviews were transcribed. The interviews help illuminate our quantitative findings. The emphasis of our qualitative analysis is on describing how superusers in Gravity Spy used resources to support their participation and to learn the science of glitch classification. For this paper, two researchers searched through the interview transcripts and identified mentions of project resources. We recorded how the interviewees described the resources, how it supported learning, and at which point in their lifespan the resources were used.

5 RESULTS
The quantitative data contained the interactions of 10,732 volunteers who visited the Gravity Spy project over one year and nine months. We analyzed 2,680,830 project interactions: 2,523,670 (94%) of those interactions were classifications and the remaining 157,160 interactions were event data. Both the classification and event data are highly skewed with many volunteers contributing infrequently
or in small amounts and a handful of volunteers generating the majority of interactions in the project — a common characteristic of participation in open collaboration platforms. While the average volunteer contributed during 4.71 ($\sigma = 29.94$) sessions, the median number of sessions is one. We also find similar skewed distributions when we aggregated the number of interactions. The average volunteer generated 249.79 ($\sigma = 1472.25$) interactions, however, the median is 44. Additionally, since most volunteers contribute in only one session, a large portion of the interactions are classifications ($\mu = 235.15$, $\tilde{x} = 2$, $\sigma = 1278.03$) while a smaller number of volunteers’ interactions were non-classify event data such as opening the project’s field guide and reading discussion threads ($\mu = 4.64 \tilde{x} = 2$, $\sigma = 238.19$).

5.1 Volunteer Performance

Only 9,327 (87%) volunteers classified long enough to see gold data. Volunteers were exposed to 591,889 gold images. The mean number of gold data analyzed by each volunteer is 63.45 ($\tilde{x} = 22$, $\sigma = 171.62$). When we assessed volunteers’ historical performance, we found that volunteers performed well overall. The static average performance, not controlling for glitch, workflow or volunteers’ experience was 94.5% ($\tilde{x} = 98$, $\sigma = 0.09$).

Figure 2 depicts the average performance of all volunteers during the first 200 gold classifications (Note that the y-axis is truncated at 95%). One factor contributing to the slight decline in volunteers’ performance shown in Figure 2 could that as volunteers are promoted to more challenging workflows, they are asked to classify more kind of glitches, with lower performance on the more challenging types of glitches and more difficulty when distinguishing among a large set of options.

Figure 3 visualizes the average performance by glitch type over the first 30 exposures to each of the twenty known glitch categories. The trend lines represented in Figure 3 reveal that performance is impacted by the particular glitch category, as certain glitches classes can be learned more easily than others. For instance, the mean performance for the wandering line glitch at volunteers’ first exposure is only 47% ($N = 653$ classifications) while blip begins at nearly 100% ($N = 9,025$ classifications). After the tenth exposure to the wandering line glitch, the performance increases to 66%. Additionally, the glitches with a plateaued accuracy line are in Workflow 1 (i.e., blip and whistle) while glitches with rising accuracy trajectories (e.g., wandering line and helix) are in more advanced workflows. Further analysis of the accuracy trend lines shows that some glitch classes have a gradual increase in performance (e.g., wandering line) while others have more abrupt increases (e.g., helix).
5.2 H1a, H2a, and H3a: Resource Use and Temporality

We examine volunteers’ interactions with the system over time by grouping them into sessions. Figure 4 shows a count of the different categories of resources used on average by volunteers in their first 20 sessions (smoothed to make the pattern clearer). The figure highlights the volume of authority-subject and communal resource engagements during initial participation. During their first session, volunteers collectively had 15,685 authority-subject resource engagements and 10,087 communal resource engagements; by the second session volunteers had almost equal numbers of authority-subject (N = 4,796) and communal (N = 4,788) engagements (the totals drop as volunteers do not continue with the project).

Communal resources become a focal point of volunteers’ resource use by session three. Communal forms of presence consist of resource engagements involving other volunteers. Communal forms of presence may enroll two groups of actions – passive and active. We define passive types of communal presence as engaging in reading comments posted by other volunteers. In contrast,
active forms of communal presence appear as volunteers creating content to be consumed by other volunteers. As an example, the discussion threads are a medium to receive feedback about the answer supplied to the system since no information about performance is provided to volunteers for non-gold classifications. Some volunteers pose questions to solicit feedback from other volunteers. In interviews, Britney noted, "discussing it with other people can help introduce a new understanding of those objects too because maybe there is like a confluence of glitches, and they all take place at the same time". These discussions are also mutually reinforcing since, once produced on the system, can support other volunteers’ learning as references to a glitch’s morphological features and their location in the image.

Agent-centered resource engagements were a smaller but not insignificant type of engagement during session one (N = 2,819); however, by session 10, the number of engagements was similar to authority-subject. Given the uncertainty associated with supplying answers to glitches, some volunteers may favor images or append them to collections in order to retrieve them later as they work to reconcile the morphological features defining the glitch. For instance, one volunteer created a collection titled, “106-virgo-02a-koi-fish” which is a collection of images, “Similar to LIGO Koi Fish, glitches resemble a koi fish with fins coming out of a triangular- or teardrop-shaped body. Unlike LIGO Koi Fish though, they tend to start at slightly lower frequencies, and tend to have a very wide and/or curvy ‘mouth’.” This type of engagement helps better define koi fish glitch for the volunteer and possibly the development of a new glitch category (which volunteers can propose) given the varied morphological features identified by the volunteer.

Early resource engagement appears to comprise not only authority-subject resources such as, e.g., opening field guide but also communal resources like accessing discussion threads. There is, however, a substantial difference in the number of agent-centered and authority-subject resources. Based on these findings, H1a is partially supported. H2a is not supported as there is no clear difference in the number of agent-centered resource engagements when compared to authority-subject. These two forms of presence appear to be equally used by volunteers. Finally, H3a is supported as the number of communal resource engagements after the initial session eclipses authority-subject.

5.3 H1b, H2b, and H3b: Effect of Different Forms of Presence

To determine how performance and resources use are linked, we examined cases where a volunteer received corrective feedback on a glitch and analyzed performance on the next glitch of the same class. We checked if learning resources of the different types had been used before the next classification. There were 32,053 instances of corrective feedback received and 18,916 cases where a volunteer saw the same glitch they answered incorrectly in the same session; of these 14,283 (76%) were accurate and 4,633 (24%) were inaccurate.

To test our hypotheses, we estimated a mixed-effects logistic model, predicting accuracy (correct = 1, incorrect = 0) from the use of resources comprising the three forms of presence. We allowed the intercept to vary randomly by each volunteer. Table 3 shows the model results. We first modeled all responses to incorrect gold classifications. The All Workflows model (N = 18,916) shows that when a volunteer supplied an incorrect answer to a gold classification, responding accurately on the next gold classification was influenced significantly by a volunteer having used agent-centered and authority-subject resources. According to the model, using an agent-centered resource engagements is associated with 0.337 (p = 0.005) higher log odds when compared to those who have no agent-centered resource engagements. Additionally, using authority-subject resources is associated with a 0.249 (p = 0.004) higher log odds of responding correctly.

Given that performance is different for different glitches and Workflows include different glitch classes, we analyzed each Workflow independently. Examining resource engagement by Workflow
Dependent variable: Accuracy (correct = 1, incorrect = 0)

(All Workflows) (Workflow 1) (Workflow 2) (Workflow 3) (Workflow 4)

<table>
<thead>
<tr>
<th></th>
<th>Agent–Centered</th>
<th>Authority–Subject</th>
<th>Communal</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.337**</td>
<td>0.249**</td>
<td>0.162</td>
<td>2.005***</td>
</tr>
<tr>
<td></td>
<td>(0.122)</td>
<td>(0.087)</td>
<td>(0.086)</td>
<td>(0.052)</td>
</tr>
<tr>
<td></td>
<td>1.690</td>
<td>0.609</td>
<td>-0.359</td>
<td>5.291***</td>
</tr>
<tr>
<td></td>
<td>(0.963)</td>
<td>(0.348)</td>
<td>(0.447)</td>
<td>(0.490)</td>
</tr>
<tr>
<td></td>
<td>0.476</td>
<td>0.557*</td>
<td>0.371</td>
<td>1.969***</td>
</tr>
<tr>
<td></td>
<td>(0.430)</td>
<td>(0.224)</td>
<td>(0.251)</td>
<td>(0.115)</td>
</tr>
<tr>
<td></td>
<td>0.466</td>
<td>0.614*</td>
<td>0.379</td>
<td>2.203***</td>
</tr>
<tr>
<td></td>
<td>(0.449)</td>
<td>(0.312)</td>
<td>(0.295)</td>
<td>(0.143)</td>
</tr>
<tr>
<td></td>
<td>0.513**</td>
<td>0.256</td>
<td>0.304*</td>
<td>0.964***</td>
</tr>
<tr>
<td></td>
<td>(0.195)</td>
<td>(0.150)</td>
<td>(0.147)</td>
<td>(0.077)</td>
</tr>
</tbody>
</table>

Observations: 18,916 2,973 2,899 2,490 4,458
Log Likelihood: -9,394.887 -1,198.839 -1,385.204 -1,080.509 -2,523.552
Akaike Inf. Crit.: 18,799.770 2,407.678 2,780.408 2,171.018 5,057.104
Bayesian Inf. Crit.: 18,839.010 2,437.665 2,810.268 2,200.118 5,089.117

Note: *p<0.05; **p<0.01; ***p<0.001

Table 3. The results of mixed-effects logistic regression estimation by workflow. The dependent variable is volunteer response on the next glitch of the same type. The random effect is user.

is important since as volunteers are promoted to higher workflows, there are more glitches making the task more difficult for volunteers. Further, progress through the Workflow levels is an indication of the volunteer’s development and engagement with the system.

While there are many resource engagement during volunteers’ interactions in Workflow 1, the model shows that no resource engagement types significantly influence the likelihood of a correct response. Workflows 2 and 3 introduces volunteers to more classes of glitches. The results of the mixed-effects model in Table 3 show that having authority-subject resource engagements in Workflow 2 increases the log odds of responding correctly by 0.557 ($p = 0.013$) and in Workflow 2 by 0.614 ($p = 0.049$). The results indicate the usefulness of authority-subject resources in these early stages of engagement in the project.

This quantitative finding is supported by our qualitative data. In interviews with superusers, they often pointed to the field guide and tutorial as important for their initial learning the intricacies of the glitch categories. Britney described her early participation and the resources that were valuable as she learned to become a contributor. Britney first described the field guide and how the resource provided information about glitch identification saying, “when I was first starting I think I looked at them [field guide] pretty frequently just to make sure I was getting it right and understanding what I was looking at.” For Britney the field guide was an authoritative source from which she could retrieve information to help her performance in identifying glitches.

We observed a shift in the importance of resource engagements in Workflow 4 in Table 3; using authority-centred resources is longer a significant predictor of performance, and agent-centered and communal resources instead become predictive. Workflow 4 shows that having authority-subject resource engagements is associated with a 0.513 ($p = 0.008$) higher log odds of answering accurately.
Again, these findings are supported by interview data in which volunteers describe their participation practices. Audrey discussed her use of collections (an agent-centred resource), stating, “After I see a certain pattern a few times I create a new collection and search in collections of other volunteers so that my collection name match with possible others.” By sifting through the many data objects and identifying prototypical glitch examples, Audrey creates resources for other volunteers to consult. She will annotate these collections with comments describing the glitch features, which later helps newcomers recognize confusing glitches. Jason also became heavily involved in curating glitches. He did so by writing a browsing script (to the surprise of the Zooniverse software developers) to page through many glitch images. His program scrapes through the URLs on the site, which allows him to examine them closely for interesting features at his own leisure.

Communal-focused resource engagement such as posting comments and viewing discussion threads become important for improving performance (Workflow 4 in Table 3). Having communal resource engagements is associated 0.304 (p = 0.038) higher log odds of answering accurately. Interviews with superusers point to the importance of viewing the talk pages to pose questions, receive feedback, and converse with other members. Britney’s work centered around providing what she described as feedback on the collections of other volunteers. She described her approach, saying she would leave comments like “hey this is interesting although it does not quite fit, here is what it kind of looks like but here is the reason why it does not quite fit I think and then maybe #possiblenewglitch.”

H1b, H2b, and H3b were derived from volunteers’ efforts to improve their performance after having responded incorrectly on a glitch classification. The models presented in Table 3 reveal that for Workflows 2 and 3, agent-centered resource engagements are important for improving their ability to classify glitches and for classifying in Workflow 4 agent-centered and communal resources are most useful, supporting our hypotheses.

6 DISCUSSION

6.1 Learning Resources to Improve Performance

Our results show that on average Gravity Spy volunteers are accurate classifiers; however accuracy is mediated several factors including workflow and volunteers’ engagement with learning resources. When examined from Sørensen’s forms of presence we observed shifts in the types of learning resources used to aid learning. H1a posits that during early stages volunteers will seek authority-subject resource engagements to improve learning. The visual inspection of resources used over-time show a preference for authority-subject resources in initial stages, however, the use of authority-subject resources during later stages appear to decrease. The results of our mixed-effects logistic regression show that the use of these resources is associated with improved performance, at least in Workflows 2 and 3. We suspect this was not the case in Workflow 1 since the two glitch categories were specifically chosen for this level for their obvious distinguishing characteristics to make them easy to learn without support, facilitating newcomers to join the project. The glitch categories in Workflows 2 and 3 are more challenging than those presented in Workflow 1 and have increased inter-category variability. Thus, volunteers improve learning by engaging with resources constructed by the project organizers, i.e., the field guide and tutorials to help learn the classes.

While engagement with agent-centered resources are crucial to early learning, we observed shifts in the types of presence volunteers seek during more advanced stages in their participation, as agent-centered and communal interactions (H2a and H3a) become more used (see Figure 4). These shifts are also crucial to volunteer learning (H2b and H3b). We posit two reasons for this change. First, we believe that by this point in their engagement with the task, volunteers have largely absorbed the static authority-subject learning materials, making it less useful to consult...
them again. Furthermore, the categorization task in Workflow 4 is more challenging than the previous workflows. The glitches added at this level show more variability, making them harder to classify, and there are more choices, some with subtle distinctions, making the overall classification task harder. The variability is hard to capture in a few instances, meaning that the various authority-subject resources may also be less useful. Through interviews with superusers and our analysis, we found that volunteers rely on individual organizing (agent-centered) and social participation (communal) resource engagements to reconcile learning under uncertainty.

Learning with an emphasis on individual organizing (i.e., agent-centered) constitutes volunteers working independently to resolve learning gaps. Based on interviews with superusers, we see increased interactions curating content for one’s self. Since each volunteer’s collections and favorites can be accessed through their personalized homepage, they can build a repository of glitch images to define features of the glitch to help resolve the variability. This individual organizing serves to help advanced learners better grapple with the variety of glitch classes and document different glitch morphologies. When power users interact on the site, the content they generate serves two purposes, one intentional and the other unintended. As power users curate glitch images, placing them in collections and marking them as favorites, they intentionally aggregate resources to better support their participation in identifying new glitch classes. An unintended consequence of this work is the trail of resources left for other volunteers to consume. A common issue noted by volunteers is the lack of prototypical examples for the classification task, especially for newly identified classes. As well, there is a lack of guidance that covers marginal examples. As power volunteers collect and debate where these glitches fall, newcomers often approach these as resources from which to learn.

In the process of building collections, volunteers also engage in conversations about the shape and form of glitches. Social participation evolves as volunteers engage with other volunteers. We can make a distinction between reading and contributing to the social spaces of the project. The variation in less-certain glitch classes can be learned through discussions with other volunteers and science team members. Because volunteers cannot view other volunteers’ classifications, they must instead rely on what Mugar et al. [19] called practice proxies, that is, discussions that convey how the task is performed. We find that for more advanced volunteers, the discussion forums are not simply a place to observe practice, but also contribute to practice. This importance charts comparing resource use from interviews with superusers reveal a shift from observing practice to creating practice. This shift represents the difference between simply viewing the discussion to adding to it. This interplay supports what Østerlund et al. [23] describes as different forms of social presence where volunteers’ relationships to resources change as they gain experience and become more knowledgeable about the task and the community.

### 6.2 Supporting Learning

Our research contributes to the e-learning literature with a deeper understanding of how user-generated resources contribute to learning in online communities. Jones and de Laat [15] argue that learner-centric ecologies involve connections (1) among learners, (2) between learners and tutors and (3) between learners and resources. We find comparable relations where learners connect through forum posts. They interact with experts (tutors or teachers) through feedback and access resources on the site. In Gravity Spy, our results suggest that resources play an essential role when direct interactions among members and experts are limited; this is particularly true early in volunteers’ tenure with a project. As volunteers become better, they tend to move to social and collaborative learning spaces. Initially, resources produced by authorities hold prominence and are replaced by access to user-generated (i.e., agent-centered and communal) content. We think that this situation may apply more generally.
The forms of presence may depend on the stage of engagement. Mugar et al. [19] noted that newcomers approach traces of other participants’ work on the discussion boards as authoritative resources as opposed to communal resources since they have little knowledge about the glitch categories. Additionally, users may seek resources outside the interactions supplied on the project website. One superuser, Mike, noted he consulted journal articles outside the site in addition to information on the site to help him understand the science behind gravitational wave research. Information found in the about pages. Mike suggested these resources are important since “…I don’t know of anywhere else I can go on, say the LIGO collaboration website and find relevant data.”

On peer-production platforms, users might encounter a few formal tutorials and readings, but a lot of the attention falls on user-generated content developed through engagement with the platform. For instance, on Wikipedia, users can find resources on how to make their first contribution or how to becomes an administrator, both curated by the community of Wikipedians. These distinctions suggest in the e-learning literature; more attention is given to how learning resources are assembled to support learning, interactions among participants, and relations to the experts.

### 6.3 Implications for Designing Online Communities

Given the empirical results, the question of how best to organize resources on the site is important. Connectivism [5], which suggests learning cannot be designed; it can only be “designed for” by creating infrastructures that allow individuals to make connections to the online environment. Our results highlight movement from the individual to social and collaborative learning resources. These findings suggest a different view on Luckin [17] learner-centric ecologies. She argues that resources should be actively organized and administered. Our findings suggest that this organization needs to extend to the resources created by users. However, the question of how best to organize user-generated content and to integrate it with the formal training materials for newcomers remains an open question. In this case, the resources become a part of the ecosystem of learning materials but are not easily found. Indeed, the difficulty of navigating the large volume of talk may be the reason for the importance of the search function.

Based on these findings, we suggest that online community managers continually evaluate resources and direct users to new resources that support learning. We also suggest for online production communities; site managers begin to enroll the resources generated by the community by referencing the materials created by users in the authoritative resources such as the projects about pages or the field guide. This evaluation supports additional research on how users use, create, and assemble resources.

### 6.4 Limitations

As with any research project, there are limitations. The main limitation of our data analysis is that it is based on what the system records in the system logs. There is a tacit assumption that clicking on resources are indicative of engagement with that resource. However, we do not measure how long volunteers stayed on the page or whether they even read the text. Additionally, the system does not collect the URLs of specific pages, which could have allowed us to draw more fine-grained conclusions about the resources learners use. For example, the analysis might reveal more specifically which discussion threads are helpful.

A second limitation is that we only examined the resources which are available or generated in the Gravity Spy project. We know from interviews that some users consult resources external to the project site, e.g., scientific publications and videos, which might help users to have a better understanding of gravitational waves and glitch detection. However, we have no way to track the use of such resources. A third limitation is that we removed sessions where users had no gold data. These are likely sessions in which users had a small number of classifications; nevertheless, the
activities of these users could be important in efforts to draw conclusions about the resource used to support learning in Gravity Spy.

Finally, Gravity Spy allows volunteers to contribute anonymously, and there are 183,338 anonymous records contained in this dataset. These anonymous records are potentially important since an important part of the learning process is a volunteer’s initial interactions with the project reading materials like the project’s About pages or posts made by other volunteers. The initial exploration may be important for learning; however, they are not captured by the system. Research by Jackson et al. [12] found that 14% and 7% volunteers in Higgs Hunters Gravity Spy respectively contribute anonymously and in comparing distributions of classification behaviors, datasets which include anonymous traces led to statistically different data distributions.

ACKNOWLEDGMENTS

Without the Zooniverse volunteers who worked on the projects, there would be no paper. Many thanks to the Zooniverse team for access to data. This project was partially supported by the National Science Foundation, Grant No. INSPIRE 15-47880.

REFERENCES


