

# Shifting Forms of Engagement: Volunteer Learning in Online Citizen Science

COREY BRIAN JACKSON, University of California, Berkeley, USA and Syracuse University, USA

CARSTEN ØSTERLUND, Syracuse University, USA

MAHBOOBEH HARANDI, Syracuse University, USA

KEVIN CROWSTON, Syracuse University, USA

LAURA TROUILLE, Adler Planetarium, USA

Peer production projects involve people in many tasks, from editing articles to analyzing datasets. To facilitate mastery of these practices, projects 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. We draw on Sørensen's framework of forms of presence to distinguish three types of engagement with learning resources: authoritative, agent-centered and communal. We assigned resources from the Gravity Spy citizen-science into these three categories and analyzed trace data recording interactions with resources 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 peer production.

CCS Concepts: • **Human-centered computing** → **Computer supported cooperative work**.

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

## ACM Reference Format:

Corey Brian Jackson, Carsten østerlund, Mahboobeh Harandi, Kevin Crowston, and Laura Trouille. 2020. Shifting Forms of Engagement: Volunteer Learning in Online Citizen Science. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 36 (May 2020), 19 pages. <https://doi.org/10.1145/3392841>

## 1 INTRODUCTION

Learning by doing characterizes many peer production projects, such as Wikipedia, open-source software, or citizen science. Even so, projects typically provide at least some resources to train and socialize new project 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 [9]. StackOverflow, a Q&A community, provides new users with a

---

Authors' addresses: Corey Brian Jackson, coreyb.jackson@berkeley.edu, University of California, Berkeley, USA, 102 South Hall, Berkeley, California, 94720; Carsten østerlund, Carsten.østerlund@syr.edu, Syracuse University, USA, 343 Hinds Hall, Syracuse, New York, 13210; Mahboobeh Harandi, mharandi@syr.edu, Syracuse University, USA, 343 Hinds Hall, Syracuse, New York, 13210; Kevin Crowston, crowston@syr.edu, Syracuse University, USA, 343 Hinds Hall, Syracuse, New York, 13210; Laura Trouille, trouille@zooiniverse.org, Adler Planetarium, USA, Chicago, Illinois, 60605.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2020 Association for Computing Machinery.

2573-0142/2020/5-ART36 \$15.00

<https://doi.org/10.1145/3392841>

two-minute tutorial covering how best to formulate questions and the benefits of applying tags. Projects might also advertise best practices in frequently asked questions (FAQs) or about pages. Mugar et al. [20] found that in Planet Hunters, a virtual citizen science project, the comments posted by experienced users serve as valuable learning resources for newcomers, as the comments pointed to specific features of work practice that were lacking in project tutorials.

FAQ, how-to pages, and comments may be valuable resources for learning and socialization, but we know little about how members of the community make use of them over the course of their tenure with 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 projects 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 at the beginning stages of participation but lose its significance over time.

In this article, we draw on digital 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 amateur volunteers in scientific tasks ranging from selecting research questions to publishing scientific results [2]. 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 learning resources to educate the volunteers. While there are several models of citizen science and a range of tasks in which to involve volunteers, the project we investigated is conducted exclusively online and asks volunteers to classify objects in pre-existing data.

Specifically, we present an analysis of digital trace representing the resources used by volunteers in Gravity Spy, a project hosted on the Zooniverse platform [27]. Gravity Spy records volunteers' accuracy on classification task, providing the authors an opportunity to assess how well 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 (e.g., newcomer) affect performance. Specifically, we address the question: **Which project resource engagements best support learning at different stages of engagement?**

## 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 on such crowdsourced and peer-based platforms. Many studies have a strong educational focus [7, 11, 12, 16, 18, 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 conceived as activity systems [8], communities of practice [17], sociocultural [25], or socio-material formations [28]. 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 interactions with other participants and experts. Engagement with peer production project resources support participants' learning and thus their ability to do the required work.

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 participation highlights the importance of learners gradually gaining access to more and more communal practices [17].

The zone of proximal development concept emphasizes the need to adjust learning opportunities to the trajectory of specific learners [8]. 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 [15]. A study of Wikipedia [3] shows, for instance, how novices often start by simply reading other's articles before making their initial contributions. Likewise, Preece and Shneiderman [24] suggests that participants in peer-production sites move from readers to leaders. In both instance, people gradually gain access to more involved tasks.

To characterize kinds of engagement with learning resources appropriate at different stages, we turn to work by Sørensen [28]. Her practice-based approach 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

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 prototypical knowledge displayed in the authority's region. These forms of authoritative knowledge will often be chunked up and presented to the learner in carefully organized sequences. Resources originate from the authoritative region where experts and organizers typically produce them.

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 [19], Willett et al. [32] applied examples of expert solutions of micro-tasks to train newcomers and calibrate their performances. Several scholars have found that expert feedback or prototypical examples curated by experts can aid learning. Oleson et al. [21] introduced gold standard data in a programmatic manner to train newcomers, transmitting authoritative practices in the process. Likewise, Bateman et al. [1], Spohrer and Soloway [29] and Walker and Engel [30] highlight the importance of expert feedback. In the context of citizen science specifically, researchers have found that expert and project organizer-supplied resources (e.g., FAQ, field guides) are valuable promoters of learning [10, 14, 23].

Overall, authority-subject forms of presence and associated resources appear effective early in the tenure of a peer-production participant. Authoritative resources will typically originate from experts and project organizers. 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 produced by experts 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:** Early participation in the project will contain more authority-subject engagements than agent-centered or communal engagements.
- **H1b:** Authority-subject engagements will improve task performance early in a volunteer's tenure with the project.

## 2.2 Agent-centered form of presence

Fluid relations characterize agent-centered presence and learning. Individual agency directs participants' activities, as if you let a group of 4th graders loose on Minecraft [28]. 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, but rather by the practice. As such, it moves beyond the stable relations found in authority-subject form of presence and can be hard to scaffold. It is an aggregate process that evolves over time, where a participant gradually build knowledge by adding resources to the infrastructure. The fluid learning associated with agent-centered presence involves ongoing mutations of the resources 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. In short, resources associated with agent-centered forms of presence will typically be compiled by individual learners or assist them in their knowledge search.

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 [4]. She searched for similar examples to build a collection of similar images and resources that might explain this phenomenon. Later, her agent-centered learning inspired other forms of presence where community members and experts helped in the search. Eventually, it led to the publication of an article describing this previously unknown astronomical phenomenon.

We hypothesize that resources associated with agent-centered forms of presence become prevalent once a volunteer has learned the basics of how to navigate and contribute to a project. In other words, we expect that more advanced participants will engage agent-centered resources (e.g., user-created collections of images and search features) 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:** Later participation in the project will contain more agent-centered engagements than authority-subject engagements.
- **H2b:** Agent-centered engagements will improve task performance later in a volunteer's tenure with the project.

## 2.3 Communal form of presence

Communal presence forms around collective activities, objects and events. Learning associated with communal presence takes place as participants join the community and strengthen their relations to it. Knowledge becomes validated through joint engagement in the communal practice, for instance, 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. Resources constructed by other volunteers and shared with the whole community typically support this form of presence.

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 latter, contributions in a communal setting [6, 34]. Doroudi et al. [6] 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, the authors found that reviewing peer-work solutions was effective perhaps because workers had to actively engage with the work of their peers, rather than simply taking them as given [5].

Weir et al. [31] coin the concept of learner-sourcing to describe the inclusion of learning materials sourced from learners (or the crowd). 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 [20]. To newcomers these traces act as proxies for advanced 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 to e.g., discussion boards and other communal features can thus serve as user-generated learning resources, accessed as part of participation in the community. Thus, we hypothesize that:

- **H3a:** Later participation in the project will contain more communal engagements than authority-subject engagements.
- **H3b:** Communal engagements will improve task performance later in a volunteer's tenure with the project.

At this point, we are left uncertain about the relative importance of agent-centered and communal forms of presence and resource engagement. Neither Sørensen [28] nor the peer-production literature offer any clear hints to whether one dominates the other or if they both increase their importance in tandem as learners 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 [33], 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 also results 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 (bottom) shows the classification interface.

**Scaffolded Participation.** A novel feature of Gravity Spy is an implementation of a training regime that gradually introduces glitch classes to volunteers: first two, then five, and only after demonstrating sufficient competence in identifying glitches, all of the glitch classes [33]. As a result, participation in Gravity Spy is scaffolded: volunteers start in Workflow 1 and based on their

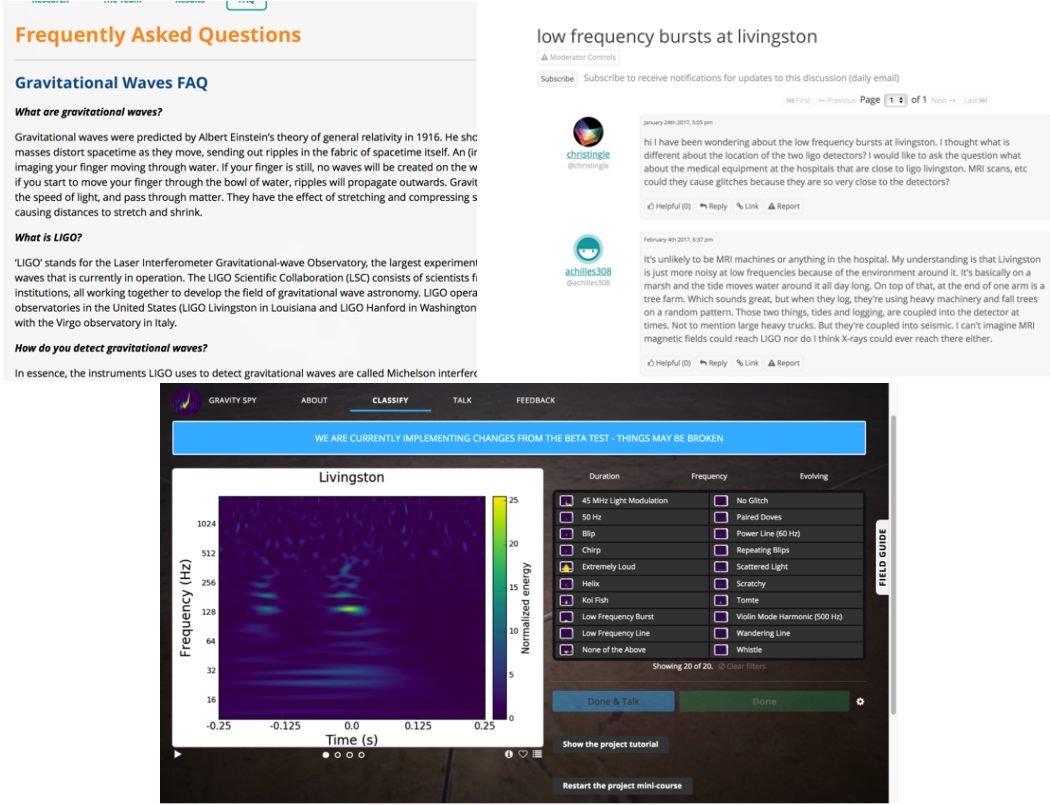


Fig. 1. Screenshots of Gravity Spy pages. The FAQ page (top, left) provides volunteers with answers common questions posed by volunteers such as “What is LIGO?”, an example of a discussion thread (top, right) in the project where a volunteer posed a question about a glitch at an observation location, and the classification interface (bottom). Volunteers are presented with a spectrogram and asked to categorize the noise signal represented in image.

performance can advance to upper levels. Table 1 shows each workflow and the new glitch classes at each level in bold font.

Importantly, 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. 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 which reads,

*“You responded [Volunteer’s Selection]. When our experts classified this image, they labeled it as a [Correct Choice]. 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 system applies a crowdsourcing classifier 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. As soon as a volunteer joins the project and starts the classification, the system makes a confusion matrix for them. The matrix is updated when the volunteer classifies an image from the gold standard data and also when an image is retired. The reliability model is built based on the



Workflow (# New Glitches)	Glitches available
Workflow 1 (2)	<b>blip, whistle</b>
Workflow 2 (3)	blip, whistle, <b>koi fish, power line, violin mode</b>
Workflow 3 (4)	blip, whistle, koi fish, power line, violin mode, <b>chirp, low frequency burst, no glitch, scattered light</b>
Workflow 4 (10)	blip, whistle, koi fish, power line, violin mode, chirp, low frequency burst, no glitch, scattered light, <b>helix, 45Mhz light modulation, low frequency noise fluctuations, paired doves, 50hz, repeating blips, scratchy, tomte, wandering line, extremely loud</b>

Table 1. The glitches available to volunteers in each workflow (or level). The glitches in bold are new in that level.

probability of providing a true label by the volunteer for each category. Volunteers are promoted if their reliability vector passes the reliability threshold vector for each category.

**Learning Engagements on Gravity Spy.** The Gravity Spy project site contains a number of learning resources developed by experts, individual volunteers and the community (i.e., authority-subject, agent-centred and communal forms of presence). As examples of authority-subject forms of presence, 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 prototype displayed to volunteers.

As examples of agent-centred forms of presence, the site supports search functionality and collections for individual 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”<sup>1</sup> to investigate glitches that, as stated by the volunteer, “share the same 0.4 Hz timing as paireddoves, but with a different morphology and with a weak amplitude.”

As examples of communal forms of presence, the project hosts five 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 (top, right) 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 deliberations what features of an image causes it to be classified as one glitch category versus the other, volunteers sometimes leave

<sup>1</sup><https://www.zooniverse.org/projects/zooniverse/gravity-spy/talk/729/207235>

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 about contributions or for noting bugs in the interface. The discussion board is a general board for conversations about any topic.

## 4 METHODS

Our study is a part of a larger research project developing a citizen science project with human-machine collaboration components. To address our research questions and hypotheses, we apply a mixed-method approach combining quantitative data in the form of digital traces of volunteers interactions on the Gravity Spy website and qualitative data in the form of volunteer interviews.

### 4.1 Data Collection

Digital trace data span one year and nine months beginning in March 2016. We acquired two forms of trace data: (1) interaction data representing the click behaviors (e.g., clicking the FAQ page) and (2) event data representing the classification submissions. Each interaction contained a userID (the unique identifier for a volunteer), the URL path of the click (e.g., *about/viewdiscussion*), and a timestamp. Classification records contained a subjectID (the unique identifier for an image), a userID, an indication conveying whether the subject was gold standard, the volunteer’s classification response, and a timestamp.

To prepare the trace data for analysis, we executed several data processing steps. First, the URL paths generated by the system contains the provenance of the volunteer’s interactions leading to a web-page. For example, when a volunteer accessed a discussion thread while on the classify page, the URL path was recorded as */classify/viewdiscussion*, however, accessing the same discussion thread from the Gravity Spy homepage recorded */about/viewdiscussion* as the URL path. We clustered together URL paths that led to the same part of the system to obtain a list of types of system pages that can be visited.

Next, two authors collaboratively coded each type of page to determine whether the page was a learning resource and its form of presence as described by Sørensen. Our approach to determining whether a page would be included as a learning resource was to simply exclude those that led to pages without textual data. For instance, we excluded the Gravity Spy landing page (*/home/gravityspy*). We identified 43 types of pages as having potential to serve as learning resource engagements.

Using Sørensen’s forms of presence as a theoretical lens, we coded the pages into one of the three categories. While Sørensen describes a broad range of practices around resources, our grouping was determined by who produced the resource. If scientists produced the resource, we considered it authority-subject ( $N = 3$ ); if the volunteer produced the resource, we considered it agent-centered ( $N = 7$ ); and if a community of volunteers produced the resource, we considered it communal ( $N = 33$ ).<sup>2</sup> Table 2 lists examples of typical resources associated with each form of presence. Note that our records are of the clicks leading to particular pages. Some kinds of resources are more complex than others, and so leave multiple traces, e.g., building a collection click-by-click, item-by-item.

Next, the coded interaction data were combined with classification, which allowed the authors to construct a detailed traces of volunteers engagements on Gravity Spy. Finally, to assess relative engagement with different kinds of resources we clustered the trace data into larger units.

<sup>2</sup>Part of the reason for the large number of communal pages is that interactions with the different discussion board are coded separately.



Specifically, traces were aggregated at the session level, a sequence of interactions carried out in single sitting. As volunteers do not always log out of the system when they are done classifying, we determine the end of a session when the gap to the following interaction was greater than 30 minutes. We derived a session variable for each record in the trace data indicating to which session a series of events belonged.

Form	# Resources Engaged - Examples
Authority-Subject	3 - field-guide, tutorial, image metadata
Agent-Centered	7 - user-collections, user-favorites, search
Communal	33 - talk-board, member-profile, post comment

Table 2. Typical resources associated with each form of presence.

While our main analysis is quantitative, to help in interpreting the data, we conducted four semi-structured interviews with superusers to understand their practices doing advanced work in Gravity Spy. The best-performing volunteers are promoted to the highest workflow, having demonstrated classification mastery of gold standard data being exposed to all twenty glitches with varied levels of difficulty. Superusers' experiences span a multitude of practices including accessing resources and socializing with other volunteers. The interviews helped to define the context surrounding some resource engagements and their relative importance overtime. The interview participants we interviewed were active across 372 sessions and executed 9,794 classifications. We recruited 10 interview participants through private messages sent through the Gravity Spy website. Four volunteers responded to our interview requests. Each interview was conducted over the telephone and lasted approximately one hour and were transcribed.

## 4.2 Data Analysis

To address H1a, H2a, and H3a, we used the Mann-Kendall test to detect significant trends in resource engagement. The Mann-Kendall is a non-parametric trend test that assesses whether there is a monotonic upward or downward trend in an independent variable. To address our hypotheses, a session over session time series was created with resource engagements in each form of presence as the independent variable. Given the high variance in resource engagements, analysis using the Mann-Kendall are not impacted by the magnitude of extreme values.

H1b, H2b, and H3b seeks to determine whether some forms of presence are more salient for learning than other resource engagements. To assess learning, we examined resource engagements and performance after a volunteer mis-classified a gold-labelled glitch. We extracted all mis-classified glitches and the type of glitch. We then identified the next occasion that a volunteer saw a gold-labeled glitch of the same type (i.e., the follow-up glitch). We then identified whether a volunteer had resource engagements in each of the three forms of presence (1 = yes, 0 = no) between the two gold-labeled glitches. We used a mixed-effects logistic regression to determine the significance of each form of presence on improving performance for the *follow-up glitch response* (1 = correct classification, 0 = incorrect classification). A mixed-effects logistic regression model is appropriate because it allows us to control for the non-independence of data at the user level since volunteers classify multiple images. The model included main effects variables for the three forms of presence as binary variables indicating whether a volunteer engaged resources between making the two classifications. The equation:

$$\text{follow-up glitch response} \sim \text{authority-subject} + \text{agent-centered} + \text{communal} + 1/\text{userID}$$

We carried out this analysis on the entire data set and also by workflow to assess how the use of resources changes with experience on the site.

Since participation statistics in citizen science projects are often skewed where a handful of volunteers are responsible for a large volume of contributions, reporting the mean in descriptive statistics may misrepresent common behaviors [26]. Thus, in addition to reporting the mean ( $\mu$ ) and standard deviations ( $\sigma$ ), we also report on the median values using the notation  $\tilde{x}$ . All corresponding data analysis and visualizations were computed using the RStudio.

The interviews were used to provide context to resource engagements. Two authors read through the interview transcripts and searched interview notes to identify responses where interviewees mentioned the resource engagements listed in Table 2. In total, we identified eleven statements where volunteers mentioned project resources and interactions that contributed to learning. As we discussed and synthesized the quantitative results, the interview data was used to provide the missing context to the use of resource engagements in volunteers' learning processes. We use the pseudonyms Britney, Mike, Jason and Audrey when referring to interviewees.

## 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 digital traces: 2,523,670 (94%) were classifications and 157,160 (6%) were click 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 peer-production projects. 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 had 249.79 ( $\sigma = 1472.25$ ) interactions, however, the median is 44. A large portion of the interactions are classifications ( $\mu = 235.15$ ,  $\sigma = 1278.03$ ,  $\tilde{x} = 2$ ) and a smaller number of volunteers' interactions included event data such as opening the project's field guide and reading discussion threads ( $\mu = 4.64$ ,  $\sigma = 238.19$ ,  $\tilde{x} = 2$ ).

### 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 ( $\sigma = 171.62$ ,  $\tilde{x} = 22$ ). 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% ( $\sigma = 0.09$ ,  $\tilde{x} = 98$ ).

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 is 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

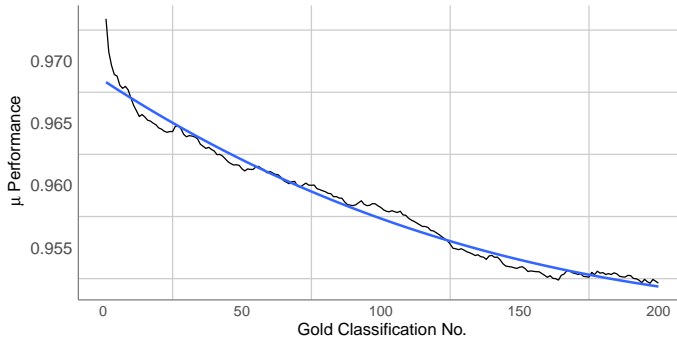


Fig. 2. The mean performance over the first 200 gold classifications.

classes have a gradual increase in performance (e.g., wandering line) while others have more abrupt increases (e.g., helix).

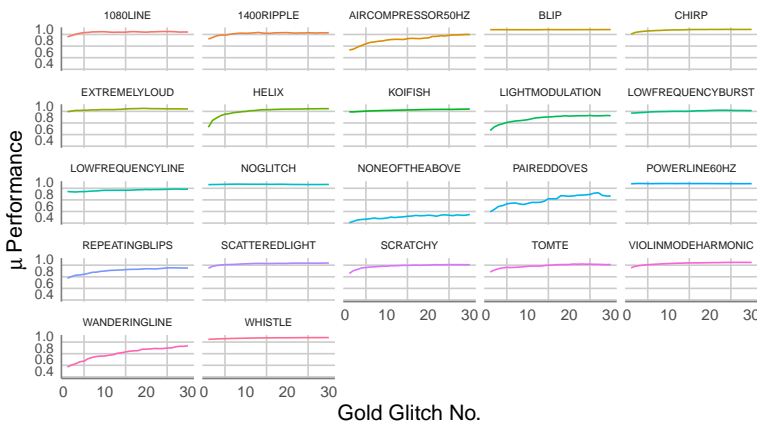


Fig. 3. The mean performance over the first 30 gold standard classifications for each glitch category.

## 5.2 H1a, H2a, and H3a: Resource Use and Temporality

Figure 4 shows the mean number of resource engagements across twenty sessions for each form of presence. The most frequent form of presence during volunteers' first session was authority-subject ( $N = 15,685$ ,  $\mu = 1.46$ ) resource engagements followed by communal ( $N = 10,087$ ,  $\mu = 0.93$ ) and then agent-centered ( $N = 2,819$ ,  $\mu = 0.26$ ). However, by the second session engagement with authority-subject and communal resource engagements have equal means, at  $\mu = 1.5$ . Agent-centered engagement remains relatively low at  $\mu = 0.36$ . Based on visual inspection of Figure 4, it is quite evident that authority-subject resources engagements decline while agent-centered and communal engagements increase. The Mann-Kendall trend test showed a significant downward trend for authority-subject resource engagements ( $z = -5.09$ ,  $p < 0.001$ ).

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

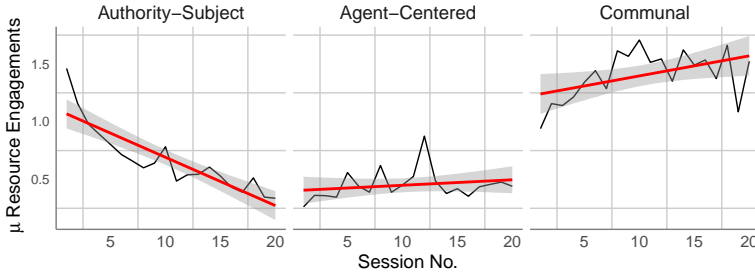


Fig. 4. The  $\mu$  number of interactions for each form of presence over the first 20 sessions with a fitted regression line.

volunteers may favorite images or append them to collections in order to retrieve them later as they work to reconcile the morphological features defining the glitch, engaging in agent-centered forms of presence. 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. We observed an upward trend in resource engagements, but it was not statistically significant ( $z = 1.4$ ,  $p = 0.163$ ).

Communal engagements showed a significant upward trend ( $z = 2.11$ ,  $p = 0.038$ ). Communal forms of presence consist of resource engagements involving other volunteers. 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*”.

Early resource engagement appears to comprise not only authority-subject forms of presence but also communal forms of presence. 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: Effects of Different Forms of Presence

To determine how performance and learning practices 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 counted the interaction types and grouped them into the three forms of presence. 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, a necessity to account for the non-independence of the data as volunteers can have many classification records. 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, when using authority-subject resources, the odds of answering correctly increase by a factor of 1.283 ( $p = 0.004$ ). The marginal effect of non-use is .88 while use of authority-subject resources increases the marginal effect to .91. Additionally, when using an agent-centered resource engagements the odds of answering correctly increase by a factor of 1.401 ( $p = 0.005$ ) compared to those who have no agent-centered resource engagements. The marginal effects are the same as authority-subject resources (.88 for non-use and .91 for use).

	<i>Dependent variable:</i>				
	follow-up response (correct =1, incorrect = 0)				
	All Workflows	Workflow 1	Workflow 2	Workflow 3	Workflow 4
Authority-Subject	1.283**	1.838	1.745*	1.848*	1.291
Agent-Centered	1.401**	5.421	1.610	1.594	1.671**
Communal	1.176	0.698	1.448	1.460	1.355*
Constant	7.424***	198.566***	7.166***	9.052***	2.623***
Unique Volunteers	4,137	1,832	1,652	1,156	732
Observations	18,916	2,973	2,899	2,490	4,458
Log Likelihood	-9,394.89	-1,198.84	-1,385.22	-1,080.52	-2,523.55
Akaike Inf. Crit.	18,799.77	2,407.68	2,780.41	2,171.02	5,057.10
Bayesian Inf. Crit.	18,839.01	2,437.67	2,810.27	2,200.12	5,089.12

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 volunteer.

Given that performance is different for different glitches and Workflows include different glitch classes, we analyzed each Workflow independently. Examining resource engagement by Workflow also shows how use of resources changes with experience, since progress through the Workflow levels is an indication of the volunteer's development and engagement with the system. Furthermore, as volunteers are promoted to higher workflows, there are more glitches making the task more challenging for volunteers.

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. We believe that this null result reflects the fact that Workflow 1 has only 2 relatively easy classes of glitch, meaning that there is not much to learn. 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 odds of responding correctly by 1.745 ( $p = 0.013$ ) with marginal effects for non-use at .88, compare to .93 in cases where authority-subject resource engagements occurred. In Workflow 3 we observed, a 1.848 ( $p = 0.049$ ) increase in odds when authority-subject resource engagements occur and use results in a marginal

effect of .95 while volunteers who do not have authority-subjects can be expected to be accurate 90 percent of the time. 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.

In Workflow 4, We observed a shift in the importance of resource engagements; using authority-centred resources is no longer a significant predictor of performance, and agent-centered and communal resources instead become predictive. Workflow 4 shows that the odds of being correct increase by a factor of 1.671 ( $p = 0.008$ ) when agent-centered resource engagements occur with marginal effects of use at .83, while volunteers who do not have agent-centered resource engagements are only .74 likely to respond accurately. Communal resource engagements increase accuracy by a factor of 1.355 ( $p = 0.038$ ) and the marginal effects associated with engagements was .79 compared to .74 when a volunteer does not have communal engagements.

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 also become important for improving performance in Workflow 4. 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, authority-subject 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

Our results show that on average Gravity Spy volunteers are accurate classifiers; however accuracy is mediated by several factors including workflow and volunteers’ resource engagements.



### 6.1 Learning Resources to Improve Performance

Using Sørensen to examine 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. As we stated, 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. The marginal probabilities reported in section 5.3 reveal the magnitude of improvements in cases where a volunteer engages with a particular resource. We observed a 5 to 8 percent improvement in the likelihood of answering accurately across all workflows. This improvement, while small, has important implications for volunteers and the system. For volunteers, during early workflows, an individual is promoted five times faster, and during later workflows, using communal and agent-centered resources means a volunteer is promoted eight times more quickly. For the system, beyond achieving higher data quality, resulting in faster image retirement leading to more rapid analysis of the data.

While engagement with authority-subject 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. Second, the categorization task in Workflow 4 is more challenging than the previous workflows. The glitches added at this level show more variability, introducing more choices and requiring subtle distinctions, all of which make the classification task harder. A few examples cannot capture this diversity, meaning that the authority-subject resources may also be less useful. Through our trace data analysis and interviews with superusers, we found that volunteers rely on agent-centered and communal resource engagements to reconcile learning under uncertainty.

Volunteers working independently to resolve a learning gap involve searches for and organizing of resources (i.e., agent-centered). 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.

Volunteers also engage in conversations about the shape and form of glitches (i.e., communal presence). 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.

## 6.2 Enhancing Learning in Peer Production

Given the empirical results, the question of how best to organize resources on the site is important. Our results highlight movement from the individual to social and collaborative learning resource engagements. When power users interact on the site, content they generate serves one primary and possible one secondary purpose. 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. The development of communal resources evolves as volunteers engage with other volunteers.

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 become 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.

The forms of presence may depend on the stage of engagement. Mugar et al. [20] 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 project's 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."*

Based on these findings, we suggest that project organizers continually evaluate resources and direct users to new resources that support learning. We also suggest that projects leverage the resources generated by the community by referencing these user-created materials in the authoritative resources such as the project's about pages or the field guide. This evaluation supports additional research on how users use, create, and assemble resources.

## 7 LIMITATIONS AND FUTURE WORK

The study comes with a number of limitations associated with our trace data methodology. First, the system logs set limits to our data analysis. We assume that clicking on resources indicate engagement with that resource. However, we do not measure how long volunteers stayed on the page or whether they even read the text. The traces serve as atomized proxies of people's interactions and we do not know the depth of each resource engagements. A quick peek at the FAQ requires less effort compared to posting a detailed comment. 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.

Second, Sørensen's theory about forms of presence encompasses a broad range of practices that involve both the production and use of resources. In this study, we operationalize the resource engagement quite narrowly according to who has produced the resource. The Gravity Spy trace data

allowed us to unambiguously determine who had produced the resource (i.e., authority, individual, community). We did not have data that would allow us to classify resources engagement according to how they were used in context. Such an approach would likely lead to a richer understanding of resource engagement where one resource type might support different forms of presence in different uses. Interviews reveal a shift from observing practice to creating practice or the difference between viewing discussions and contributing to discussions. Future research should seek to develop measures that captures not only who produce a resource but also how it is used and for how long. The way that volunteers combine resources in a session might hint at their use. For instance, just looking at talk posts following each classification might hint at an authority-subject form of presence while the combination of talk and discussion post may suggest communal presence.

Third, we only examined the resources available or generated in the Gravity Spy project. We know from interviews that a few 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. Finally, several studies note the limitations of trace data as often being incomplete accounts of people's interactions as some platforms allow users to participate without having an account [13]. In this dataset, volunteers can participate anonymously only during Workflow 1. While the possibility of anonymity exists, the anonymous traces we excluded originate in Workflow 1, and do not affect our findings at higher Workflow levels. Future work will determine how to include anonymous resource engagements.

## 8 CONCLUSION

In this article, we investigate the behaviors of volunteers in an online citizen science project. Applying Sørensen's theory about forms of presence, we determine how volunteers' resource engagements change as their relationship to the project changes and the complexity of the task increases. Volunteers begin their tenure as citizen scientists relying primarily on authority-subject and communal engagements. As the task difficulty increases, volunteers' reliance on authority-subject engagements subsides significantly and other forms of engagements become more central namely agent-centered and communal.

These results are buttressed by responses from superusers who provide context to the quantitative findings presented in this article. Volunteers emphasized that, as the glitch classification task becomes more challenging, social participation becomes necessary as a form of question asking, learning about the practices of other volunteers, and performing their own analysis to make sense of ambiguous glitch classes. Overall, the findings suggests that social participation and independent work can enhance the learning experiences of participants in peer production. Project organizers should consider approaches to include communal resources in people's learning experiences and provide directed instruction for independent work.

## ACKNOWLEDGMENTS

Without the Zooniverse volunteers who worked on the projects, there would be no article. Many thanks to the Zooniverse team for access to data and our research collaborators: Sarah Allen, Sara Bahaadini, Scotty Coughlin, Vicky Kalogera, Aggelos Katsaggelos, Shane Larson, Neda Rohani, Joshua Smith, and Michael Zevin. This project was partially supported by the National Science Foundation, Grant No. INSPIRE 15-47880.

## REFERENCES

- [1] Scott Bateman, Jaime Teevan, and Ryen W White. 2012. The search dashboard: how reflection and comparison impact search behavior. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012), 1785–1794.
- [2] Rick Bonney, Heidi Ballard, Rebecca Jordan, Ellen McCallie, Tina Phillips, Jennifer Shirk, and Candie C Wilderman. 2009. *Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report*. Technical Report.
- [3] Susan L Bryant, Andrea Forte, and Amy S Bruckman. 2005. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*. ACM, 1–10.
- [4] Carolin Cardamone, Kevin Schawinski, Marc Sarzi, Steven P Bamford, Nicola Bennert, C M Urry, Chris Lintott, William C Keel, John Parejko, Robert C Nichol, Daniel Thomas, Dan Andreescu, Phil Murray, Jordan Raddick, Anže Slosar, Alex Szalay, and Jan Vandenberg. 2009. Galaxy Zoo Green Peas: discovery of a class of compact extremely star-forming galaxies. *Monthly Notices of the Royal Astronomical Society* 399, 3 (Nov. 2009), 1191–1205.
- [5] Shayan Doroudi, Ece Kamar, and Emma Brunskill. 2019. Not Everyone Writes Good Examples but Good Examples Can Come from Anywhere. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 17.
- [6] Shayan Doroudi, Ece Kamar, Emma Brunskill, and Eric Horvitz. 2016. Toward a Learning Science for Complex Crowdsourcing Tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2623–2634.
- [7] Stephen Downes. 2006. Learning networks and connective knowledge. *Collective intelligence and elearning* 20 (2006), 1–26.
- [8] Yrjö Engeström. 2014. *Learning by expanding*. Cambridge University Press.
- [9] Aaron Halfaker, Aniket Kittur, and John Riedl. 2011. Don't Bite the Newbies: How Reverts Affect the Quantity and Quality of Wikipedia Work. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*. Association for Computing Machinery, New York, NY, USA, 163–172. <https://doi.org/10.1145/2038558.2038585>
- [10] Katie DeVries Hassman, Gabriel Mugar, Carsten Østerlund, and Corey Jackson. 2013. Learning at the Seafloor, Looking at the Sky: The Relationship Between Individual Tasks and Collaborative Engagement in Two Citizen Science Projects. In *Proceedings of the 10th International Conference on Computer Supported Collaborative Learning (CSCL)*.
- [11] Caroline Haythornthwaite. 2014. New Media, New Literacies, and New Forms of Learning. *International Journal of Learning and Media* 4, 3-4 (2014), 1–8.
- [12] Caroline A Haythornthwaite and Richard Andrews. 2011. *E-learning Theory & Practice*. Los Angeles : Sage.
- [13] Corey Brian Jackson, Kevin Crowston, and Carsten Østerlund. 2018. Did they login? Patterns of Anonymous Contributions in Online Communities. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–16.
- [14] Corey Brian Jackson, Carsten Østerlund, Gabriel Mugar, Katie DeVries Hassman, and Kevin Crowston. 2014. Motivations for Sustained Participation in Crowdsourcing: Case Studies of Citizen Science on the Role of Talk. In *2015 48th Hawaii International Conference on System Sciences (HICSS)*. IEEE, 1624–1634.
- [15] Aditya Johri and Seungwon Yang. 2017. Scaffolded Help for Learning: How Experts Collaboratively Support Newcomer Participation in Online Communities. In *Proceedings of the 8th International Conference on Communities and Technologies*. ACM, New York, NY, USA, 149–158.
- [16] Chris Jones and Maarten de Laat. 2016. Network Learning. In *The SAGE Handbook of E-learning Research*, C Haythornthwaite, R Andrews, J Fransman, and E M Meyers (Eds.). 44–61.
- [17] Jean Lave and Etienne Wenger. 1991. *Situated Learning. Legitimate Peripheral Participation*. Cambridge University Press., NY.
- [18] Rosemary Luckin. 2008. The learner centric ecology of resources: A framework for using technology to scaffold learning. *Computers & Education* 50, 2 (Feb. 2008), 449–462.
- [19] Tanushree Mitra and Eric Gilbert. 2014. The language that gets people to give. In *Proceedings of the ACM 2014 conference on Computer Supported Cooperative Work*. ACM Press, New York, New York, USA, 49–61.
- [20] Gabriel Mugar, Carsten Østerlund, Katie DeVries Hassman, Kevin Crowston, and Corey Brian Jackson. 2014. Planet hunters and seafloor explorers: legitimate peripheral participation through practice proxies in online citizen science. *Proceedings of the ACM 2014 conference on Computer Supported Cooperative Work* (2014), 109–119.
- [21] David Oleson, Alexander Sorokin, Greg Laughlin, Vaughn Hester, John Le, and Lukas Biewald. 2011. Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing. *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence* (Aug. 2011).
- [22] Carsten Østerlund and Paul Carlile. 2005. Relations in Practice: Sorting Through Practice Theories on Knowledge Sharing in Complex Organizations. *The Information Society* 21, 2 (April 2005), 91–107.
- [23] Carsten Østerlund, Gabriel Mugar, Corey Brian Jackson, Katie DeVries Hassman, and Kevin Crowston. 2014. Socializing the Crowd: Learning to talk in citizen science. In *Academy of Management Annual Meeting, OCIS Division*. Philadelphia, PA.

- [24] Jennifer Preece and Ben Shneiderman. 2009. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction* 1, 1 (March 2009), 13–32.
- [25] Brian J Reiser. 2002. Why scaffolding should sometimes make tasks more difficult for learners. In *Proceedings of the conference on computer support for collaborative learning: Foundations for a CSCL community*. International Society of the Learning Sciences, 255–264.
- [26] Henry Sauermann and Chiara Franzoni. 2015. Crowd science user contribution patterns and their implications. *Proceedings of the national academy of sciences* 112, 3 (2015), 679–684.
- [27] Robert Simpson, Kevin R. Page, and David De Roure. 2014. Zooniverse: Observing the World’s Largest Citizen Science Platform. In *Proceedings of the 23rd International Conference on World Wide Web (WWW ’14 Companion)*. ACM, New York, NY, USA, 1049–1054. <https://doi.org/10.1145/2567948.2579215>
- [28] Estrid Sørensen. 2009. *The Materiality of Learning*. Cambridge University Press.
- [29] James C Spohrer and Elliot Soloway. 1985. Putting It All Together Is Hard for Novice Programmers. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics* (1985).
- [30] Henry M Walker and Kevin Engel. 2006. Research exercises: immersion experiences to promote information literacy. *Journal of Computing Sciences in Colleges* 21, 4 (April 2006), 61–68.
- [31] Sarah Weir, Juho Kim, Krzysztof Z. Gajos, and Robert C. Miller. 2015. Learnersourcing Subgoal Labels for How-to Videos. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW ’15)*. Association for Computing Machinery, New York, NY, USA, 405–416. <https://doi.org/10.1145/2675133.2675219>
- [32] Wesley Willett, Jeffrey Heer, and Maneesh Agrawala. 2012. Strategies for crowdsourcing social data analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 227–236.
- [33] Michael Zevin, Scott Coughlin, Sara Bahaadini, Emre Besler, Neda Rohani, Sarah Allen, Miriam Cabero, Kevin Crowston, Aggelos K Katsaggelos, Shane L Larson, Tae Kyoung Lee, Chris Lintott, Tyson B Littenberg, Andrew Lundgren, Carsten Øesterlund, Joshua R Smith, Laura Trouille, and Vicky Kalogera. 2016. Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science. *arXiv* 6 (Nov. 2016), 064003. arXiv:1611.04596
- [34] Haiyi Zhu, Steven P Dow, Robert E Kraut, and Aniket Kittur. 2014. Reviewing versus doing. In *Proceedings of the ACM 2014 conference on Computer Supported Cooperative Work*. ACM Press, New York, New York, USA, 1445–1455.

Received October 2019; revised January 2020; accepted March 2020