

# Planet Hunters and Seafloor Explorers: Legitimate Peripheral Participation Through Practice Proxies in Online Citizen Science

Gabriel Mugar, Carsten Østerlund, Katie DeVries Hassman,  
Kevin Crowston, Corey Brian Jackson

Syracuse University, School of Information Studies  
{gmugar, costerlu, klhassma, crowston, cjacks04}@syr.edu

## ABSTRACT

Making visible the process of user participation in online crowd sourced initiatives has been shown to help new users understand the norms of participation [2]. However, in many settings, participants lack full access to others' work. Merging the theory of legitimate peripheral participation [18] with Erickson and Kellogg's theory of social translucence [10, 11, 16] we introduce the concept of practice proxies: traces of user participation in online environment that act as resources to orient newcomers towards the norms of practice. Through a combination of virtual [14] and trace ethnography [12] we explore how new users in two online citizen science projects engage with these traces of practice as a way of compensating for a lack of access to the process of the work itself. Our findings suggest that newcomers seek out practice proxies in the social features of the projects that highlight contextualized and specific characteristics of primary work practice.

## Author Keywords

Guides; instructions;

## ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces - Interaction styles.

## General Terms

Human Factors; Design; Measurement.

## INTRODUCTION

A critical issue for sustaining groups that need to persist over time is how newcomers to the group learn to be effective participants. In some groups, new members go through formal educational or orientation activities in order to learn group practices. However, researchers have argued that formal education alone does not convey the necessary tacit knowledge about work practices needed for good performance. Such tacit knowledge can be conveyed instead through informal learning experiences, such as the process of legitimate peripheral participation (LPP), which

describes modes of situated learning whereby newcomers start off by engaging in simple practices and observing more experienced members of a community as they engage in their work practices [18].

Online groups often face difficulties around newcomer orientation. Many online groups are composed of members who are not part of a single formal organization and who contribute only in their free time, reducing or eliminating the possibility of formal training. However, the affordance of the technology used to support group interaction often make it possible for distributed volunteers to observe work in progress, thus enabling a form of LPP. For example, Bryant et al. [2], studied new Wikipedia participants and suggested that new editors began by reading articles before they make their initial contributions.

In this paper, we examine newcomer learning in a kind of online citizen science project: non-temporary groups in which large numbers of distributed volunteers collaborate with domain scientists to analyze large data sets to fulfill scientific goals. These projects are an intriguing example of distributed learning and knowledge production supported by public engagement in scientific research processes [15, 26]. Specifically, we examined two projects developed as part of the Zooniverse<sup>1</sup>: Planet Hunters (PH) and Seafloor Explorer (SE), in which members of the general public are asked to annotate scientific data and photographs of phenomena of interest (evidence of possible planets and marine organisms, respectively). To be effective over time, the projects must facilitate new users to orient themselves towards the goals and work practice of the project. However, unlike other online projects like Wikipedia, PH and SE participants are not able to see the work other users have done, in this case, the primary annotations they have made. The scientific task was deliberately designed with this restriction to ensure independent responses by eliminating the possibility that one user's response to an image could affect responses from other users.

Given that project participants can not observe the results of the primary work practices of other participants, we pose the question of if and how informal experiential learning

---

<sup>1</sup> <https://www.zooniverse.org/>

such as LPP might work in such a setting. To address this question, we extend Erickson and Kellogg's work on social translucence and social proxies [10, 11, 16] to consider "practice proxies." We define practice proxies as system features that make visible the socially salient aspects of people's unfolding work practices rather than the practices directly. In doing so, we maintain Erickson and Kellogg's focus on design features in online environments that allow newcomers to observe traces of others activities, but emphasize the translucence of work practices as opposed to social norms. By reflecting on the design implications of the project as they relate to LPP and access to practice proxies, we ask how characteristics of practice proxies support new users given the lack of access to traces of primary project work practices.

We conclude by reflecting on how our analysis of practice proxies in Zooniverse applies to other open online collaborative communities. While deliberately limited access to other people's work in Zooniverse is a unique characteristic of this project, setting it apart from many open online collaborative communities, we find this design condition to be a revelatory opportunity for exploring practice proxies and the ways in which access to observing practice is made possible in online collaborative environments. On further reflection, it is clear that even in settings where the work product are shared, such as free/libre open source software (FLOSS) development or Wikipedia, many details of the work practices remain private (e.g., design, testing and debugging in FLOSS). Through our analysis, we will demonstrate the ways in which practice proxies in Zooniverse, which appear primarily in discussions about work, provide access to practice in the same ways talk page conversations and edit histories in Wikipedia, or release notes and bug reports in FLOSS projects, or conversations between Xerox machine repair men [22], provide access to practice.

## THEORY

The concept of legitimate peripheral participation (LPP) describes the process of moving from being a newcomer and outsider to becoming an insider to a set of practices, in a community of practice. Lave and Wenger [15] coined the term to describe learning among apprentices in a range of different fields from tailors in Liberia, Mayan midwives in the Yucatan, U.S. navy quartermasters, non-drinking alcoholics, and U.S. supermarket meat cutters. Newcomers start out by participating in a practice, or set of practices, in a way that makes them legitimate but peripheral members of the community. Socially, they move towards the center of the community, as they increasingly become sustained participants fluent in the tasks, vocabulary and organizational principles of the community.

Legitimate peripheral participation articulates two types of practices that newcomers gradually gain legitimate access

to: 1) their own participation and 2) the participation of others. First, as indicated by the concept, newcomers gradually gain access to participate in the practices of the community. The new apprentices described in Lave and Wenger's study start out engaging in low-risk practices. The community can afford if they fail. Gradually they gain access to more elaborate forms of participation. For example, apprentice tailors start out detailing the nearly finished garment such as sewing on buttons. Following the production process in reverse they gradually move from detailing, to sewing to finally cutting cloth.

Second, LPP highlights the role of newcomers' access to other people's practices. If novices can observe more experienced participants in their daily work they can develop an understanding of the context and the various activities, process and activities central to becoming a sustained participant. Lave and Wenger offer an iconic counter-example of apprentice butchers in a supermarket where the physical layout of the space does not give them access to the work of expert meat cutters. They are literally stuck in a corner engaging in menial work. There is a lack of transparency. That is, the learners have limited access to the *context* and *specificity* of other people's unfolding work [25]. To put it differently, access to the other's practices does not happen through instructions or teaching *about* the practice, but only through specific and contextualized discussions and stories told *within* practice [18]. The context only comes to life in all its specificity when newcomers experience others engaging in meaningful tasks (e.g., using the tools of the trade, bringing about coordination, negotiating disagreements or addressing uncertainties).

We find a number of studies using LPP to understand the types of participation one finds in online environments [2, 7, 23, 24]. Notably, in a study of Wikipedia, Bryant, Forte and Buckman [2] describe participant movements from newcomers to sustained users in regard to the their access to 1) participate in the community and 2) access other participants' practices. Based on interviews with nine Wikipedia participants they find that newcomers perceive no technical barriers to full participation. Neither do they articulate a lack of access to other people's practices. The very design of Wikipedia gives anybody access to the articles produced by other participants, access to discussion boards for articles, and access to the history of changes to an article.

Experiences in Wikipedia and other online communities such as open source software development (i.e., FLOSS) raise the questions: what types of transparency and thus access to other people's participation facilitates newcomers learning? In other words, what practices needs to be translucent for newcomers to learn?

There are several streams of research that addresses the ways in which participants in online collaborative environments benefit from access to work by other participants. For example, research on activity awareness (e.g. [6];[19]) describes how certain features in online collaborative environments support the work practices of a distributed collaborator base by keeping them up to date on the respective contributions of participants. In doing so, such features help collaborators make sense of each other's work, effectively tailor their respective contributions to each other's work and so better coordinate the process. Similar research on FLOSS has focused on how the visibility of completed work and the visibility of decisions behind such work acts as a critical component for coordinating FLOSS projects [5].

While the previously mentioned work focuses on the coordination of work, the literature on social translucence (e.g. [10, 20]) looks at features in online environments that specifically help participants learn normative behaviors. As with activity awareness, social translucence refers to design characteristics of online environments that help promote coherent behavior between participants by making their actions visible to each other [8-10]. However, the focus in this case, elaborates on Lave and Wenger's second type of access in LPP (i.e., transparency of other's practices), not in apprenticeship arrangements but online environments. By making the actions of participants visible, system design that provides social translucence allows people to draw on each others experience to learn to make sense of the social setting [10].

But, what part of the social does a system make translucent? As McDonald et al. explain, any interaction with a system is a case for social translucence. The question is how the architecture of the system is designed to represent such participation [20]. For instance, article edit histories in Wikipedia make some aspects of Wiki work transparent, as do the reputation systems on eBay by displaying the history of a particular member's participation. In both cases however, there are other aspects of participant contributions that may not be described that could bring more contextual salience to the work being done [20, 21].

As illustrated in Erickson's notion of social proxies, systems can also strive to make social norms translucent. Social proxies are system features that visualize the socially salient aspects of online interactions, in particular, the norms of how people behave around each other. Erickson and Kellogg explore the visualization of group behavior by creating tools that visualize the placement of individuals in a space with both placement in the space and indicators of who is speaking as proxies that provide cues regarding the norms of interaction. In Erickson's words: "By making social cues visible, and allowing visible traces to accumulate over time, we create a public resource that

allows people – especially those familiar with the interactive context – to draw inferences about what is happening that can inform the ways in which they participate, and, in turn, may ultimately shape the collective activities of the participants" [8].

However, social translucence can do more than teach norms. In Wikipedia the type of transparency newcomers experience in Bryant's et al study [2] relates not only to social proxies but also what we will term *practice proxies*: that is, system features that make visible the socially salient aspects of people's unfolding work practices. Where social proxies visualize normative expectations for interactions, practice proxies provide cues for people's unfolding practices by illuminating other people's unfolding work activities through online traces. This perspective leads us back to the questions raised by LPP: what types of transparency or social illumination facilitates the movement from newcomers to sustained participants? To be more specific, what specific types of practice proxies matter for newcomers in a particular community? The practice proxies called for by newcomers in Wikipedia may not match practice proxies craved by learners in FLOSS teams or new citizen science participants. Answering these questions require that we pay careful attention to the particular communities of practices and what activities contribute to the production and reproduction of the community [17, 18]. In our study, we address these questions through an inductive study of two particular online communities, ones in which restrictions in the direct access to work products brought the role of practice proxies into clearer view.

## **METHODS**

The empirical data for this study were collected over the past nine months as part of a multi-year NSF funded action research collaboration [1], working closely with developers and designers of a collection of online citizen science projects hosted on the Zooniverse website. Each Zooniverse project (fourteen at the time of writing) is developed around a large data set provided by different science teams. The sites are designed in collaboration between scientists, web developers and educators. Our action research project strives to enhance participants' learning and motivation through system enhancement of Zooniverse's existing sociotechnical system.

### **Descriptions of Research Site**

Research and findings presented in this article focuses on two Zooniverse projects: Seafloor Explorer (SE) and Planet Hunters (PH). Both projects were developed to help scientists analyze large corpora of data by involving participants in annotating data objects. Participants in the PH project are asked to identify transiting planets in light curve images of the Cygnus constellation in order to help scientists to identify the presence of previously unknown orbiting planets. An example of the PH interface is presented in (see Figure 1). At the time of this writing, PH

participants have contributed around 17.6 million classifications that just recently contributed to the discovery of the project’s first planet. Based on analysis of participant annotation data, we find that PH contribution follows a power law distribution common to crowdsourced initiatives where many users contribute a little and a few users contribute a lot.



Figure 1. Planet Hunters annotation interface

In SE, participants are asked to identify and annotate the presence of marine specimens in images of the sea floor to help scientists better understand the species ecology of the continental shelf off the Northeastern coast of the United States (see Figure 2). Since the project was officially launched early in the Fall of 2012, SE participants have annotated over 1.4 million image objects and according to the project's blog the participants have also helped to discover a potential candidate for a new species, currently referred to as the ‘convict worm’ because of its black and white striped body.



Figure 2. Seafloor Explorer annotation interface

### Description of Data Collection and Analysis

Combining the methods of virtual ethnography [14] and trace ethnography [12] the research team collected traces of user participation in PH & SE, engaged in nine months of participant observation and conducted 10 semi-structured and focus group interviews.

First, data collection included identifying and tracing practices as they emerged based on visible participant comments, comment timestamps and other user created traces and resources within the talk and discussion features of each of the two sites. The traces left by participants’ practices (e.g., comments on talk pages) within the PH and SE online citizen science sites were used to “capture the lived experience” of participants that interact with each of the projects [12]. Trace data such as talk comments were not only analyzed as evidence of past practice, but were also analyzed as resources available to assist and guide participant engagement and coordination within the projects [12]. Participant log data collected from participant interactions and made available by the science team of each project was used to contextualize our findings. While log data may be considered as a trace of participant practice, they were not considered in the analysis of practice proxies as they are not available to participants.

Second, engaging as participant observers in both the PH & SE project enriched our understanding of participation and helped us better understand how traces of participation are left and also how they are made meaningful throughout participants’ interactions with the projects. As participants we signed up for individual user accounts and completed new participant tutorials prompted of all new users prior to their first interaction with the projects. As participant observers, conducting and completing primary annotation activities and eventually participating in the talk and discussion forum with other participants allowed us to better understand the relationships between participation and site features and social resources. Acting as participants within the studied online citizen science projects allowed us to understand the experience of being a user and also provided opportunities for reflexive analysis of user participation otherwise inaccessible through traces left on the site [14].

Finally, we conducted semi-structured interviews with project developers, members of the PH and SE science teams, moderators of the social features on the projects and PH and SE participants in order to better understand the role of the projects’ interface, systems and structures. The science teams for each project are composed of working scientists and academics in charge of the data corpus around which each of the projects are structured. Aside from managing the project, members of the science team often interact with participants on the talk forum by providing feedback on their questions. Similarly, moderators provide feedback on questions to the best of their knowledge and ensure that conversations remain civil. More peripheral PH & SE participants tend to engage in primary annotation practices and engage talk and discussion features infrequently, whereas more sustained participants may perform question asking and answering practices similar to that of the moderators. Each interview lasted approximately one hour and included questions that

addressed the perceived role of talk and discussion features for the overarching goals of the project. There were no direct questions relating to how newcomers learn to participate, however follow up questions prompted further discussions about participants' experiences as newcomers within the projects.

Using the theoretical framework of practice proxies and LPP as an analytical lens, we conducted a qualitative analysis of the interviews, participant observations and trace data. Data from the interview transcripts, participant observations and trace data were all independently analyzed by two doctoral students and then compared to identify themes relating to types of practice proxies and evidence of their use by newcomers. These emerging findings were discussed at weekly research meetings where the results from trace data, participant observation and interviews were triangulated.

## **FINDINGS**

In this section we begin by first describing the different ways in which participants can contribute to the project. We identified four modes of participation that occur similarly across the two studied projects. By outlining the different modes of observing participation, we describe the practice proxies: traces of participation visible to other users and the properties of the proxies in relationship to supporting newcomer orientation towards practice. We conclude our findings by defining three typologies of practice proxies identified in our analysis of the two projects.

Through empirical analysis we identified four modes of participation that participants engage as they move from newcomers to more sustained participants in the studied projects [13]. For this we drew on data from participant observation and an analysis of the evolution of individual participants' comment types and language use in samples drawn from both projects. Findings were triangulated with interviews with project developers, science team members, and project participants. We understand each of the four identified modes of participation to be indicative of practices that contribute to the reproduction of the PH & SE project communities.

All four of the modes are part of peripheral trajectories within the projects. We base our understanding of core and peripheral participants then on their relative position within the four identified modes of participation. Insights on peripheral participant needs and practice came from insights communicated in interviews with PH & SE participants and triangulated with content and posting patterns of participant talk and discussion comments during initial interactions with the projects. Participants' initial comments were analyzed as representative of peripheral positions. Each of the four modes of participation is discussed in more detail below.

### **Access to Participating**

The most common mode of participation within the PH & SE projects involves primary annotations of data objects. This practice is supported by the system that prompts participants with multiple questions about whether or not particular characteristics are present in the data object. In PH, for example, data is presented as a time series of points on a graph called light curves representing Kepler Telescope observations of a star (see Figure 1). Dips in the agglomeration of points may be due to transiting planets and so represents areas of interest to the scientists. Because of the noise and variability of the curves, computers are not adept at identifying patterns and aberrations in the light curves. Instead, participants are presented the images and asked to identify overall features of the light curves (e.g., variability) and dips in light curves that may indicate transiting planets. SE works in a parallel fashion, with participants identifying and annotating photographs of the sea floor for overall ground cover typology—sand, shell, gravel, cobble, boulder—and the presence of specimens from four species—sea scallops, sea stars, fish and crustaceans. Aside from annotating the presence of species, SE participants are also asked to measure each identified scallop, sea star, fish and crustacean with a measuring tool designed into the interface.

After the annotation task is completed, participants have the option to leave comments on data objects, what we refer to as user-generated annotation, and to discuss the object with other members of the community through the talk interface. User-generated annotation and asking questions about data objects are the second and third modes of participation identified, respectively. Access to queries and analysis on the talk and discussion pages of data objects can occur two ways. First, after a participant has completed a primary annotation, they are presented with an option to discuss the object further. By selecting this option, participants are brought to an object talk page where they can view annotations and queries left by past participants (see Figure 3). They are also presented with the option to start or view an ongoing discussion related to the object. The second option for accessing the talk/discussion pages for objects is by directly accessing the talk at the URL [talk.planethunters.org](http://talk.planethunters.org) or [talk.seafloorexplorer.org](http://talk.seafloorexplorer.org). There a participant can view objects that are either trending with a high number of annotations or queries, or objects that have received the most recent annotations or queries. In addition to viewing recent or trending objects, the URL for the talk pages also provides access to collections, featured discussions, recent discussions, and trending keywords.

Finally, the fourth mode of participation is participation in higher-level analysis, a practice that often exceeds the basic participation goals of the project. High-level analysis typically takes place on the talk pages of objects or on the

discussion boards. This type of analysis is often stimulated by a hypothesis or observations about data objects made by one or more participants and then communicated through the discussion forum. For example, a user might download data about a light curve and analyze it to determine characteristics of the orbit of a hypothesized planet.

### Access to Others' Participation

The traces of the first mode of participation, primary annotation, are not accessible in PH and SE. According to members of the teams that manage Zooniverse projects, the reason for this restriction is to avoid the production of bias in the annotation decisions of other participants (i.e., if participants learn in an uncontrolled manner from each other, there is a fear that they may learn and propagate incorrect data annotation practices). With the exception of primary annotations, all other practices are accessible to any participant that looks at the talk pages of data objects or the discussion forums. It is the traces of user generated annotation, queries, and higher level analysis that we define as the practice proxies of Planet Hunters and Seafloor Explorer, as they stand in for the primary work.

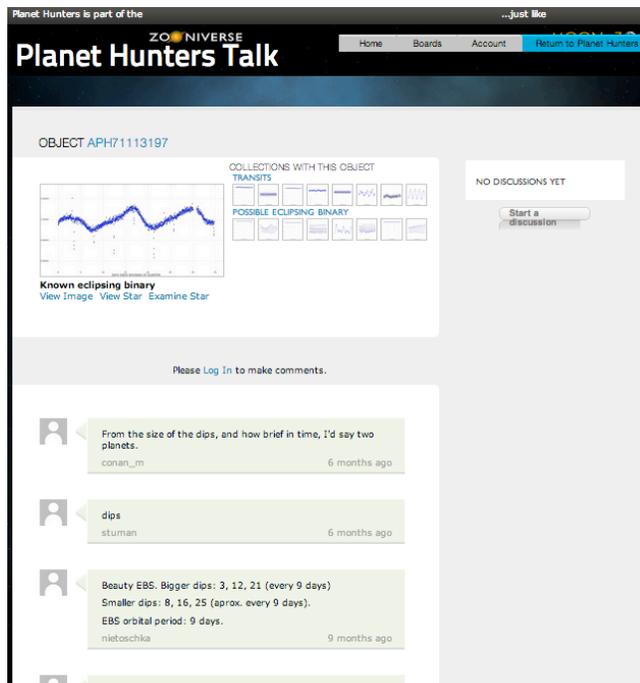


Figure 3. Talk interface for Planet Hunters.

### How Practice Proxies Perpetuate Practice

Through interviews with participants, science team members, and moderators and reflecting on our participation in the projects, we identified three themes for the ways in which new project participants interact with the practice proxies.

First, practice proxies help newcomers orient towards notable characteristics in the data. A science team member

noted that one of the benefits of the annotations and queries left by participants on the object talk pages is that they “help improve classification, because it gives you education.” In other words, the practice proxies on data object talk pages help new participants learn what characteristics in the data they should pay attention to.

Similarly, a moderator for Planet Hunters pointed out that, when they were new to the project, they actively referenced the annotations and queries left by other users on the talk pages of objects in order to learn what characteristics were important and what they looked like in the data objects.

*“New users, when they are becoming acclimated, can look at the work other users have posted and get tips on what is a transit...I know it helped me a lot when I was first doing it, to hear some of this discourse.”*

This strategy of using the traces of other users in order to learn more about the practice of the project was also used by both of our participant observers as they attempted to make sense of the practice in the project.

The second theme for evidence of use is participants referencing the work of others in order to see if they were doing their work correctly. With the option to discuss a data object after having annotated them, one of our participant observers noted that they actively looked to see if others had produced annotations or queries that matched the decisions they had made for the system annotation. One forum moderator also noted that the forums are a valuable place for new participants to see if they are doing their work correctly:

*“If there wasn't a forum, it would feel like you are doing the project on your own, you don't know if anyone else is doing it, you don't know if you are doing it right, so I think that the role of the forum is there to act as a community resource, but also to act as a backup for people when they need it.”*

An active user of PH similarly noted that early on in their engagement with the project, they actively compared the decisions they had made with the comments made on the object talk pages:

*“...most of the threads that have people posting targets to them, they are already vetting from other targets that other people found, so instead of just going to the very small, basic tutorial you get through the interface there, you can actually go check and see, “Oh this is what a bigger transit looks like, oh this is what a smaller transit looks like, oh this is what a not-transit looks like.” And just kind of figuring out, with examples if what you found is something worthy or not...”*

Additionally, we also know that the option to discuss data objects in PH is mostly used for referring to others work, as just over 69% of participants select the option, but only 22% of users that have participated leave at least one comment (See Table 1). The difference between the percentage of people who visit the object talk pages on PH but do not leave a comment may suggest that participants who click on the option to discuss mostly do so with the intention of seeing traces of other participant's practice.

Zoo	Annotation	Posts to Talk	Talk Visits	Posts to Discussion	Discussion Visits
PH	17.6M	389K	3.1 M	18K	673K
SE	1.4 M	29K	167K	511	18K

**Table 1.** Count of contributions and views in PH and SE.

The third theme describes how participants benefit from finding questions left by other participants that are similar to their own questions. In our participant observations, we found that we benefited from coming across questions generated by other participants and the responses to the questions. In the case of questions with answers, the benefit was that we were able to learn from the answers, but even in the case of those questions without answers, we benefited from the traces of other participants' questions because they indicated that our questions were in fact relevant and related to the goals of the project.

In summary, by representing the traces of other's practices, practice proxies act as a resource for newcomers to learn how to perform basic tasks in the projects. In particular, practice proxies support new users as they learn what characteristics in the data are relevant to project goals and how to identify such characteristics. Despite the presence of tutorials, interviews with science team members, moderators and project participants reveal that engaging the comments on the data object talk pages is the most valuable means of learning about relevant data characteristics and how to identify such characteristics. As one science team member stated, the tutorials showed the most ideal example of what annotating a data object might look like, which, more often than not, do not relate to the data objects that participants are likely to encounter. The situated nature of annotations and queries on the talk pages or collections allows new participants to see what parts of the data objects people looked at and what characteristics they believed to be present. Each trace of practice is thus a resource upon which new participants can draw on as they learn how to identify the presence of important characteristics in the data. Examples from the findings demonstrate practice proxies on data objects highlighting general concepts related to the project goals, specific descriptions of data characteristics, as well as questions related to conducting practice.

### How Practice Proxies Support Accurate Work

By being a resource for learning how to perform basic tasks in the projects, our findings showed us that practice proxies are also important resources for supporting accurate and correct annotation of data objects. Where some citizen science projects will approach the question of ensuring accurate work by pairing citizen scientists with trained professionals [3], the distributed and large-scale nature of projects like PH and SE does not make having a professional looking over the shoulder of a participant a feasible model. Additionally, while participants could seek out the advice of science team members on PH and SE at every turn, this would tax the science team and would not help the projects grow. The consolation that we noticed in our findings is that practice proxies act as important resources for supporting accurate and correct annotation of data objects. As our interviews revealed to us, participants often sought out the comments of other participants to see if they were doing their work correctly. In the case of projects like PH and SE, after having annotated the data object, a participant is always presented with the option to discuss the object further on the objects talk page. Here, participants can see if their choices match those of other participants based on the annotations other participants may have left. As described in the findings, this practice of checking to see if work had been done correctly was noticed in our interviews and participant observations. In an interview with a participant, the respondent pointed out that they actively sought out annotations left by more experienced users, using the traces of their practice as a resource to learn how to accurately identify the characteristics relevant to project goals.

### Practice Proxies Support Orientation of New Members

Traces of annotations and queries include participants' notes about characteristics they marked in the primary annotation practice, or comments and questions seeking clarification about the presence of particular characteristics they just engaged. In either case, such annotations and queries indicate how participants are thinking about the data objects they engage in primary annotation practice and what aspects they think are relevant to the goals of the project. As such, annotations and queries can help newcomers learn what characteristics are relevant to either primary or broader scientific objectives of the project.

In analyzing the annotations and queries, we identified three types of what we perceive to be practice proxies, or traces of user practice accessible to other users that address key aspects of normative participation in the projects. We developed the themes based on the work of Lave and Wenger's Situated Learning [18], which attends to context and specificity as conditions for supporting learning for newcomers. In developing the themes of practice proxies, we thought about how various annotations and queries lent themselves to supporting these characteristics of learning.

In this section we discuss the three types we observed and how they help primary practice. We will engage a more in depth discussion regarding their impact on context and specificity in the following section.

The first type of practice proxy draws attention to the general characteristics of the data object without making any detailed observations. For example, in PH we see annotations like *"looks like some surface activity"* or *"I also don't see evidence of planet. All downspikes are consistent with the main pulsating plot."* Similarly, in SE we see annotations like, *"Also, a funny shaped seastar,"* or *"It looks like there is something in the right bottom corner. Cannot identify it."* Such practice proxies help point out general characteristics that are important to consider when analyzing data objects, but do not make any detailed observations. By conveying the general description of characteristics that are relevant to the goals of the project, such practice proxies that draw attention to general descriptions in data objects help to establish a broad contextual framework for the goals of analyzing data in the primary practice process.

The second type of practice proxy draws attention to specific points in the data objects. For example, in PH we see one participant leaving the annotation "There appears to be a dip at day 16," and in another example, *"possible transit at 29.25."* Similarly, in SE a participant notes, *"I believe there is a sponge growing on a scallop It looks like there is a small fish, red hake, on the upper side The image quality isn't the best though to be absolutely sure of this."* In all of these examples we see traces of specific points in the data objects that represent what participants pay attention to when engaged in the primary practice of annotation. In addition to highlighting specific points in the data, such traces act as examples of the characteristics that are important to project goals.

The third practice proxy involves asking questions about the presence of characteristics in data objects. In an example from PH we see someone asking about the presence of a characteristic at a specific point in the data, *"Something at 19,"* and in another example someone appears to be questioning the presence of a phenomenon they have not seen before, *"Possibly transits at days 28,29,30, but what truncated the peak at days 6 and 7?"* As with the other examples, the overarching value of this practice proxy is that it demonstrates what aspects of the data object are important to participants and how they engage them in the process of working on the primary goals of the project. Specific to this type of proxy is that it provides insight into the questions that come up for participants as they work on the goals of the project. Having insight into the questions that other participants have about their work is something that, as we will show later in this section, helps other participants make sense of the project by finding other participants who have questions

similar to their own. Identifying similar questions can potentially help to legitimate paying attention to particular characteristics in the data, thus supporting an aspect of practice in the project.

### Context and Specificity

In our analysis of proxy types, we identified variations in the context and specificity of the practice proxies. In other words, the projects vary in how their practice proxies bring the context to life in all its specificity through discussions and stories told *within* meaningful tasks (e.g., using the tools of the trade, bring about coordination, negotiate disagreements or address uncertainties).

From the perspective of legitimate peripheral participation we understand proxies left by participants to be indicative of changing participant roles, while at the same time, due to their relative permanency with the online community, they are also then part of the reproduction of the community. Given practice proxies become part of the reproduced community as resources for new participants, a high level of detail can be important. Indeed, one moderator in the PH project related a story of telling a new volunteer of the value of being more specific when writing and creating comments. Drawing on empirical analysis we explore specificity generally, as an orienting capacity of practice proxies within online communities. We present empirical examples below and highlight proxy characteristics that impact differences in proxy specificity.



**Figure 4. Sample photo of SE project data object**

The data object, image ASF0000k9g, (see Figure 4) is one of the thousands of images presented to participants during primary annotation activities in the SE project. Scattered amongst sea scallops (orange shells), a lonely sea star, gravel, and sand are thin brown snake-like fish. The three comments below were left by SE participants in reaction to image ASF0000k9g. User B's inquiry, an example of the third type of proxy and representative of reflective practice, is directed at the presence of the snake-like creatures within the image. The description of the characteristic as "sea worms or sea snakes," adds specificity, however for

peripheral participants, this comment may be less valuable because it only specifies that the objects in question look like worms, but does not direct participants to particular locations or coordinates within the image (e.g. sea snakes in the upper left quadrant and lower half). More sustained participants however may find this level of specificity valuable as they may already be familiar with the characteristics identified in this practice proxy.

[User A] “#sand-lance #asterias #shrimp”  
 [User B] “Are these sea worms or sea snakes”  
 [User C] “These are a fish called a sand lance.”

User A’s comment is valuable as a practice proxy to new users because it is a trace of the hashtagging practice, and it also adds specificity by indicating additional data characteristics (in this example additional species names) found in the images. If a peripheral participant does not know what sand-lance, asterias and shrimp look like, they will not know which of the characteristics mentioned in comment refer to which species objects within the image. Like User B’s comment, the hashtags do not point or refer to any specific characteristic or objects within the image and thus provides a lesser amount of specificity. User C’s comment also provides additional specificity, but because his comment refers to User B’s previous comment, the specificity of User C’s comment is only valuable relative to the specificity of User B’s previous comment.

In the following example from PH we see an annotation on object APH71113197 (see Figure 5) where User A provides precise coordinates for the location of the characteristics in the data object that represent the presence of an eclipsing binary star or EBS as it is described in the talk comment below. This is an example of the second type of practice proxy, where attention is drawn to specific points in the data object, therefore providing access to specific decisions made in the practice of a participant.

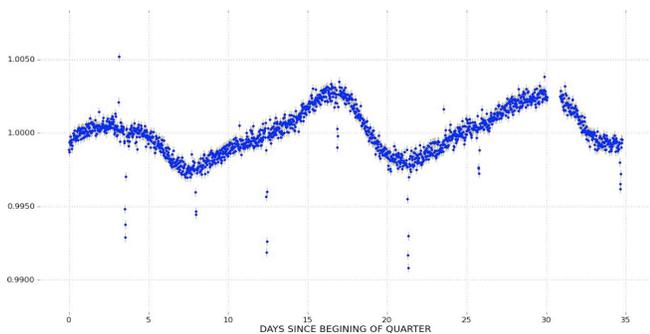


Figure 5. Planet Hunters object APH71113197.

[User A] “Beauty EBS. Bigger dips: 3, 12, 21 (every 9 days). Smaller dips: 8, 16, 25 (aprox. every 9 days). EBS orbital period: 9 days.”

Eclipsing binary stars (EBS) are defined as two stars orbiting around each other. Because each star has different level of brightness, the brightness of the two stars will register accordingly in the data object. The above annotation leaves a level of detail that helps a user understand why the characteristics in the data object represent an eclipsing binary by not only indicating the location of the dips in the light curve, but also noting that the two groups have different levels of brightness. Given the detailed justification provided and the precise coordinates noted for each of the identified characteristics, this comment has a high level of specificity. The specificity of this proxy is useful for new participants as a resource for identifying and analyzing eclipsing binaries. The specificity in this case may also be valuable to more sustained participants in the PH projects, as it provides details justifying the participant’s analysis, allowing other participants an opportunity to quickly review the comment for accuracy.

Based on our observations as well as experience as participants, we found that added levels of detail are valuable for orienting new users towards more structured understanding of the characteristics in the data object that are relevant to the goals of the project. In terms of task depth and complexity, our observation of the projects is that the level of detail required for identifying the presence of transiting planets in light curves is far greater than that required for identifying the presence of marine animals in the images of the seafloor. As such, annotations and queries in PH often necessitate a great amount of specificity and detail when pointing out the presence of a characteristic as compared to what is required in SE.

## DISCUSSION

We have found that the object-oriented discussions on PH and SE Talk Pages provide a degree of context and specificity around the descriptions of work being done that newcomers find useful as they learn how to engage in the practice of the projects.

Our findings have implications for the design of technology support for open online communities to educate newcomers to become productive participants. At first glance, it might seem that the problem of a lack of access to observe work that motivated our research is a feature only of the citizen science projects we studied. However, we argue that the same characteristic applies to a greater or lesser degree in all online communities. Taking Wikipedia as an example [2, 23], while a newcomer has access to the primary practice of other participants (article edits) as well as the edit history and talk pages for an article, not all article talk pages or edit histories provide context and specificity to describe the reasoning behind the practice of participation. The work leading to the edit could be a few seconds to correct a typo or hours of research to check a fact.

Similarly, for Free/Libre Open Source projects [4]: while other participants can see the code that is contributed, they cannot observe the work that led to that code, be it a few seconds to fix a trivial bug or hours of design and implementation for a new feature. Viewed in this light, most online communities provide only proxies for the important practices that newcomers need master.

To improve newcomer learning, design for open online communities might consider ways in which to effectively insert richer practice proxies into the experience of newcomers to help them orient themselves towards the norms of practice. Considering how the production of context and specificity via practice proxies might work in other open online communities has the potential to impact the way we conceptualize the experience of newcomers and supporting their continued participation in other open online communities of practice. The object-oriented discussion feature of Zooniverse projects such as PH and SE projects has parallels in Wikipedia talk pages oriented to a particular page or FLOSS discussions associated to particular bug reports. Designers might consider presenting lists of popular subjects that have a high frequency of practice proxies to participants at predetermined times throughout their sessions. In doing so, researchers might test the impact of presenting practice proxies to newcomers at different moments on their self-efficacy, learning and motivation. Indeed, arguably even if PH and SE were to show the products of primary annotation, the question might be whether or not simply displaying decisions, for example Wikipedia's edit history feature, would provide as rich a practice proxy as those we have described in PH and SE.

More broadly, our research contributes to the theoretical work on social translucence by extending the conversation beyond a focus on normative behavior to a focus on practice. The value of our analysis is that it introduces the idea that, for there to be online community design that promotes social translucence for the practice of participants, a practice proxy is potentially a more valuable asset for coordinating a distributed grouping of participants when it includes features that emphasize the context and specificity in the trace of a participants' practice. Simply indicating the decisions made by a participant may not be sufficient for creating a socially translucent system design. By having practice proxies that emphasize context and specificity in practice, a system has the potential to embed a richer description of the norms of practice for the community, thus helping a newcomer gain a deeper understanding of how to be an engaged and valuable member. Moving forward, we plan to engage further the question of context and specificity in the design of practice proxies. Additionally, we believe that future research would benefit from a comparison between the design of social proxies and practice proxies and their relative approaches to context and specificity.

## CONCLUSION

In this paper we took up the question of how new users orient themselves towards the goals and practices of massively distributed online collaborative projects. We framed our research with the theories of legitimate peripheral participation [18] and social proxies [8, 9] so as to explore the relationship between traces of user practice in an online environment and orientation towards participating in the project. Given the relationship between access to observing other users practice and learning how to participate in an online community, we examined two citizen science projects where the traces of primary practice are not accessible, yet new users find ways to work around this lack of access.

Our findings suggest that, while newcomers lack access to the traces of primary community practices (i.e., primary annotations of data objects), they appear to compensate for this lack of transparency by taking advantage of user-generated annotations and queries on the talk and discussion pages of data objects as a way to build their own understanding. In other words, the projects possess a rich resource of practice proxies that exist on the talk and discussion pages of the projects data objects. There, users have engaged in what we describe as user generated annotations and queries. They ask questions or make comments on the presence of particular characteristics in the data objects that are relevant to the primary annotations and broader project activities and goals. These annotations and queries are what we define as practice proxies in that they represent the traces of how users are thinking about both the primary annotations and broader project activities. The presence of such practice proxies provides resources for new users to observe as they learn how to become participants in the project. Access to the practice proxies is made available through a prominently placed link in the project interface and at the end of the projects primary practice workflow, whereby, after participants have made an annotation, they are presented with the option to view the data objects talk and discussion page.

Combining the theory of social proxy with the theory of LPP provides an analytical lens through which to observe specific design approaches for orienting new users towards the goals and practices of online crowdsourced projects. To further this particular approach to research on the design of online crowdsourced work, future research should consider design interventions that test the relationship between production and access of practice proxies to learning opportunities for new users. Building on rich qualitative and trace data such experiments might involve A/B testing the degree of access to practice proxies and measure the impact this has on continued participation.

## ACKNOWLEDGMENTS

Sample Text: We thank all the volunteers, and all publications support and staff, who wrote and provided

helpful comments on previous versions of this document. As well authors 1, 2, & 3 gratefully acknowledge the grant from NSF (grant 12-11071).

## REFERENCES

- [1] Baskerville, R. and Myers, M.D. 2004. Special Issue on Action Research in Information Systems: Making IS Research Relevant to Practice--Forward *MIS Quarterly*. 28, 3 (2004), 329–335.
- [2] Bryant, S.L. et al. 2005. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. (Nov. 2005).
- [3] Cohn, J.P. 2008. Citizen Science: Can Volunteers Do Real Research? *BioScience*. 58, 3 (2008), 192.
- [4] Crowston, K. and Fagnot, I. 2008. The motivational arc of massive virtual collaboration. (2008), 1–35.
- [5] Crowston, K. et al. 2011. Work as coordination and coordination as work: A process perspective on FLOSS development projects. (2011), 1–26.
- [6] Dourish, P. and Bellotti, V. 1992. Awareness and Coordination in Shared Workspaces. (Dec. 1992), 107–114.
- [7] Ducheneaut, N. 2005. Socialization in an Open Source Software Community: A Socio-Technical Analysis. *Computer Supported Cooperative Work (CSCW)*. 14, 4 (Jul. 2005), 323–368.
- [8] Erickson, T. 2004. Designing Online Collaborative Environments: Social Visualizations as Shared Resources. (Rutgers University, New Brunswick, NJ, Jun. 2004), 143–158.
- [9] Erickson, T. 2009. “Social” Systems: Designing Digital Systems that Support Social Intelligence. *Ai & Society*. 23, 2 (2009), 147–166.
- [10] Erickson, T. and Kellogg, W.A. 2000. Social Translucence: An Approach to Designing Systems that Support Social Processes. *ACM Transactions on Computer-Human Interaction*. 7, 1 (Mar. 2000), 59–83.
- [11] Erickson, T. et al. 2002. Social Translucence. *Communications of the ACM*.
- [12] Geiger, R.S. and Ribes, D. 2011. Trace Ethnography: Following Coordination through Documentary Practices. (2011), 1–10.
- [13] Hassman, K.D. et al. 2013. Learning at the Seafloor, Looking at the Sky: The Relationship Between Individual Tasks and Collaborative Engagement in Two Citizen Science Projects. (2013), 1–2.
- [14] Hine, C. 2000. *Virtual Ethnography*. Sage Publications.
- [15] Jason Reed, M.J.R.A.L.K.C. 2012. An Exploratory Factor Analysis of Motivations for Participating in Zooniverse, a Collection of Virtual Citizen Science Projects. (Dec. 2012), 1–10.
- [16] Kellogg, W.A. and Erickson, T. 2002. Social Translucence, Collective Awareness, and the Emergence of Place. (2002), 1–6.
- [17] Lave, J. 1988. *Cognition in Practice: Mind, mathematics, and culture in everyday life*. Cambridge University Press.
- [18] Lave, J. and Wegner, E. 1991. *Situated learning*. Cambridge University Press.
- [19] Marlow, J. and Dabbish, L. 2013. Activity Traces and Signals in Software Developer Recruitment and Hiring. (2013), 145–155.
- [20] McDonald, D.W. et al. 2013. Building for Social Translucence: A Domain Analysis and Prototype System. (2013), 637–646.
- [21] McDonald, D.W. et al. 2009. System Design for Social Translucence in Socially Mediating Technologies. (2009), 1–7.
- [22] Orr, J.E. 1986. Narratives at work: story telling as cooperative diagnostic activity. (1986), 62–72.
- [23] Preece, J. and Ben Scheiderman 2009. The Reader to Leader Framework: Motivating Technology-Mediated Social Participation. *Transactions on Human-Computer Interaction*. 1, 1 (2009), 13–32.
- [24] Schlager, M.S. and Fusco, J. 2003. Teacher Professional Development, Technology, and Communities of Practice: Are We Putting the Cart Before the Horse? *Information & Management*. 19, 3 (2003), 203–220.
- [25] Wenger, E. 1998. *Communities of Practice: Learning, meaning, and identity*. Cambridge University Press.
- [26] Wiggins, A. and Crowston, K. 2011. From Conservation to Crowdsourcing: A Typology of Citizen Science. (2011), 1–10.