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; 337 Hinds Hall, Syracuse, New York, 13244
{gmugar, costerlu, klhassma, cjacks04}@syr.edu
**National Science Foundation; 4201 Wilson Boulevard Room 1125, Arlington, Virginia, 22230
kcrowsto@nsf.gov

ABSTRACT
Making visible the process of user participation in online crowdsourced 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 environments 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
Situated Learning; Social Translucence; Legitimate Peripheral Participation; Socialization; Citizen Science

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

General Terms
Human Factors; Design; Measurement.

INTRODUCTION
To sustain a group over time, its newcomers must learn to become 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 through legitimate peripheral participation (LPP), a process of situated learning whereby newcomers initially engage in simple practices while observing the work practices and behavior of more experienced members of a community [18].

Online groups present more challenges to 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 affordances of the technologies used to support group interactions 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 begin by reading articles before they make their initial contributions.

In this paper, we examine newcomer learning in online citizen science projects: non-temporary groups in which large numbers of distributed volunteers collaborate with domain scientists to collect, annotate, identify and analyze items in 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 Zooniverse[1]: Planet Hunters (PH) and Seafloor Explorer (SE), in which members of the general public are asked to annotate scientific data (identifying evidence of possible planets and marine organisms, respectively). To be effective over time, the projects must facilitate the orientation of new users to 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.

CSCW'14, February 15 – 19, 2014, Baltimore, MD, USA.

http://dx.doi.org/10.1145/2531602.2531721

[1] https://www.zooniverse.org/

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 by other users.

Given that project participants cannot observe other participants' primary work practices, we inquire how informal experiential learning 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 aspects of 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 aspects of social translucence that relate to work practices instead of 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 their lack of access to other participants' 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 is a unique characteristic of these Zooniverse projects, setting them apart from many open online collaborative communities, this design feature offers us the opportunity to explore 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 products 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 that 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 in a community of practice of moving from being a newcomer and outsider to becoming an insider in regard to a set of practices. 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 that makes them legitimate but peripheral members of the community. Socially, they move towards the center of the community, as they become *sustained participants*, that is, active participants increasingly fluent in the tasks, vocabulary and organizational principles of the community.

Legitimate peripheral participation articulates two areas of practices in which newcomers gradually become acculturated to the community: 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 begin by engaging in low-risk practices. If they stumble and fail as they learn the norms and practices, there is little impact on the community. 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 full, 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 performing 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 conducting meaningful tasks (e.g., using the tools of the trade, bringing about coordination, negotiating disagreements or addressing uncertainties).

We find a number of studies that use LPP to understand participation in online environments [2, 7, 23, 24]. Notably, in a study of Wikipedia, Bryant, Forte and Bruckman [2] describe participant movements from newcomer to sustained user in regard to their access to 1) participate in the community and 2) to 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 need to be readily observable for newcomers to learn?

There are several streams of research that address 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 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 these studies focus 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 among 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 in 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 setting does a system make translucent? As McDonald et al. (2013) 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 are not transparent but 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, where the concept of 'socially salient' implies the prominent normative behaviors that shape and define the activities of a particular social setting.

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]

In Wikipedia the type of transparency newcomers experience in Bryant's et al study [2] relates not only to norms of interaction, but to norms of practice, or what we refer to as *practice proxies*: that is, system features that make visible the socially salient aspects of people's unfolding work practices. Where social proxies emphasize normative expectations for interactions, we explore the aspects of social translucence that act as practice proxies, providing 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 facilitate the movement from newcomer to sustained participant? To be more specific, what are the characteristics of proxies emphasizing practice that matter for newcomers in a particular community? The practice proxies called for by newcomers in Wikipedia may not match practice proxies needed by learners in FLOSS teams or new citizen science participants. Answering these questions requires that we pay careful attention to the particular communities of practice and what activities contribute to the production and reproduction of the community [17, 18]. In our study, we address these questions inductively, allowing us to explore the role of practice proxies.

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 this writing) is developed around a large data set provided by different science teams. The sites are designed in collaboration with 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

As stated above, the research and findings presented in this article focus on two Zooniverse projects: Seafloor Explorer (SE) and Planet Hunters (PH). Both projects 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 (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.

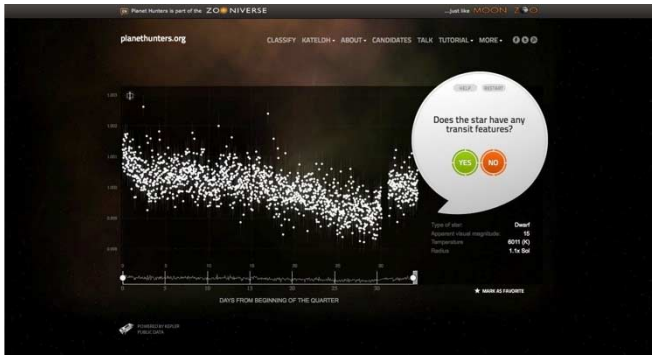

Figure 1. Planet Hunters annotation interface

In SE, participants identify and annotate the presence of marine specimens in images of the sea floor to help scientists better understand the continental shelf ecology 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.

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. Virtual and trace ethnography supplement one another. Where virtual ethnography emphasizes the participation in the online environment as well as analyzing user produced text as evidence of activity, trace ethnography focuses on participant activity as it appears in the system logs of the online platform.

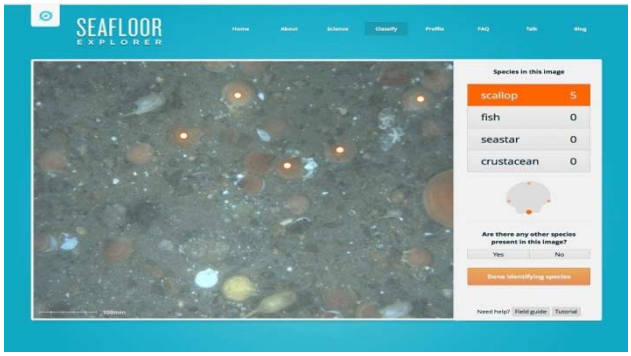

Figure 2. Seafloor Explorer annotation interface

By combining virtual and trace ethnography, researchers can identify and trace 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 [12,14]. 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 to us by Zooniverse was used to contextualize our findings. While log data may be considered traces of participant practice, we do not consider them practice proxies, as they are not made available to general Zooniverse participants.

Second, we engaged as participant observers in both the PH & SE projects, which enriched our understanding of participation, helped us better understand how traces of participation are left, and how they are made meaningful throughout participants' interactions within the projects. As participants, we signed up for individual user accounts and completed new participant tutorials that all new users are prompted to complete 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 explore the relationships between participation and site features and social resources. In short, our participation 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 in response to 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

The findings section explores the two types of practices that newcomers gradually gain legitimate access to: 1) their own participation and 2) the participation of others. First, by describing how newcomers' participation in the project and with system features changes we develop a sense of their general learning trajectory. Second, analysis of newcomers' access to fellow participants' practices opens the door to a description of practice proxies and how they promote newcomers' general learning trajectory. We outline a number of practice proxy attributes and discuss the importance of contextual detail and specificity in effective practice proxies. Finally we demonstrate how practice proxies perpetuate and support accurate citizen science work.

Access to Participating

Participants have access to four modes of participation: primary annotation, user generated annotations, user generated queries, and user generated analysis. First, the most common mode of participation within the PH & SE projects involves primary annotation 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 represent 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.

Second, 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. Third, participants can generate queries and ask questions about data objects by accessing talk/discussion pages for objects at the project website (talk.planethunters.org or 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, participants can engage 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.

The four modes of participation are also representative of a trajectory of participation that reflects a movement from the periphery to more impactful engagement with the project [13]. For example, where making primary annotations can be performed by anyone contributing to the project for the first time, making informed contributions to the Talk or Discussion pages requires a degree of understanding about project terminology and procedure [13]. High-level analysis, in the case of PH, reflects modes of contribution that can have serious implications for scientific work and discovery [13]. Indeed, it is the contribution of members

conducting high-level analysis that have led to the discovery of new planets.

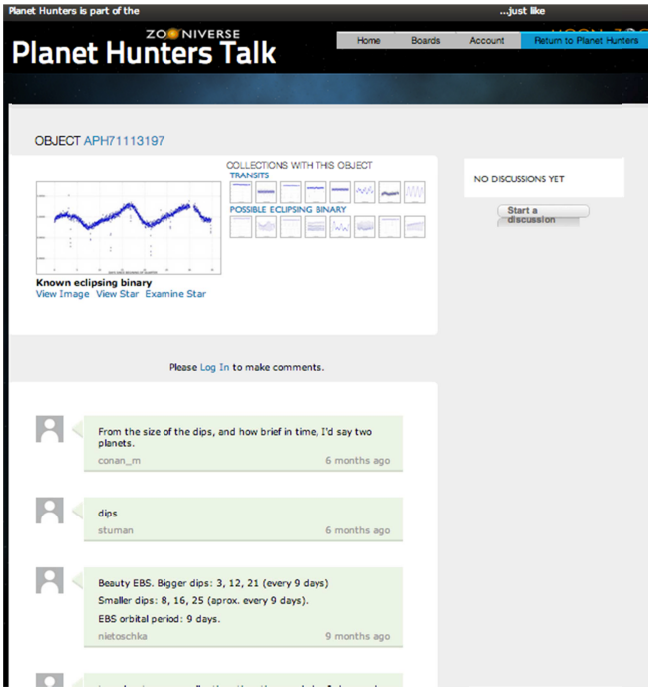

Figure 3. Talk interface for Planet Hunters

Access to Others' Participation via Practice Proxies

The traces of the first mode of participation, primary annotation, are not accessible to participants in PH and SE. According to members of the teams that manage Zooniverse projects, the reason for this restriction is to avoid bias in participants' annotation decisions (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. They can serve as a stand-in for the primary work.

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 encounter 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 doing so, practice proxies reflect resources that can support participant movement away from a peripheral position of primary annotation to more engaged positions of user generated annotation and querying.

Practice Proxy Attributes

In analyzing the annotations and queries, we identified three common practice proxy attributes. Practice proxies appear to be particular relevant to newcomers if they draw attention to 1) general characteristics of the data objects, 2) specific characteristics, or 3) ask questions about the data object. We provide examples of each feature below.

General characteristics

Some practice proxies outline 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.

Specific characteristics

Other practice proxies draw attention to specific points in the data objects. 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.

Questioning characteristics

Practice proxies also represent questions that participants have 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, these types of practice proxies demonstrate 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. Additionally, these examples reveal 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 help newcomers make sense of the project. Identifying questions similar to their own or that reflect their own confusion can potentially help newcomers to pay attention to particular characteristics in the data, thus supporting an aspect of practice in the project.

Context and Specificity

Moving beyond the specific attributes, the effectiveness of practice proxies appears to hinge on how well they articulate the context and specificity of a particular practice. For instance, 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. In other words, practice proxies vary in how well they bring the context of a practice and the specifics of the activity to life through discussions and stories told *within* meaningful tasks (e.g., using the tools of the trade, bringing about coordination, negotiating disagreements or addressing uncertainties).

Drawing on empirical analysis we explore specificity, 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.

In the SE project, the data object, image ASF0000k9g, (see Figure 4) is one of the thousands of images presented to participants during primary annotation activities. 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.

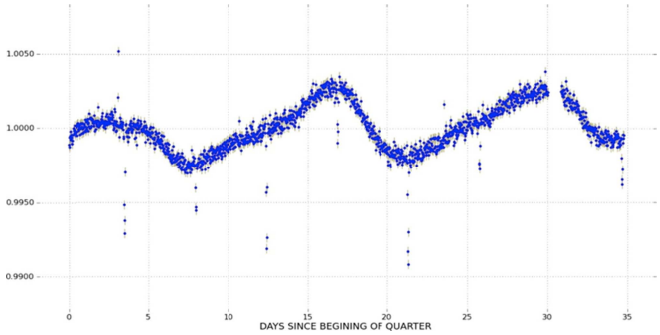

Figure 5. Planet Hunters object APH71113197

[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 levels 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 participant observation, 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. Highly contextual and specific practice proxies help participants become more skillful and engaged participants by immersing them in terminology and reasoning practices valuable for user generated annotations, queries, and higher-level analysis.

Task depth and complexity varies across projects. 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.

How Practice Proxies Perpetuate Practice

New project participants appear to interact with practice proxies in three ways. First, practice proxies help newcomers orient their own practices. A science team member noted that annotations and queries were left by participants on the object talk pages *"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.

Second, participants reference the work of others in order to assess the quality of their own work. One forum moderator 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 69% of participants access discussions, but only 22% of those actually 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

Third, participants benefit from questions left by other participants that are similar to their own questions. Traces of queries and conversations allow newcomers to map their own exploratory practice to those of others. As such they are not only seeking answers to specific questions but can attain valuable insights into the reasoning process of both fellow newcomers and more experienced participants.

How Practice Proxies Support Accurate Work

By being a resource for learning how to perform basic tasks in the projects, proxies support accurate and correct

annotation of data objects. Where some citizen science projects will ensure accurate work by pairing citizen scientists with trained professionals [3], the distributed and large-scale nature of projects like PH and SE rules out expert oversight as a feasible model. It would simply be too taxing for the science team if participants sought their feedback at every turn. Instead, practice proxies act as important resources, facilitating the accurate and correct annotation of data objects. As our interviews suggest, 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. 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 she 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.

In summary, by representing the traces of others' 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 reviewing 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 show the most ideal example of what annotating a data object might look like. More often than not, they 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 were present. Each trace of practice is thus a resource upon which new participants can draw 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.

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 acclimate to the practice of the projects. Practice proxies support newcomers' peripheral participation in the basic annotation practices and facilitate their gradual movement towards more engaged and skillful participation in the community involving user-generated annotation, queries and ultimately independent analysis. At the same time, the relative permanence of the practice proxies on the online site become an important part of the ongoing production and reproduction of the community.

Our findings have design implications for open online communities seeking to educate newcomers to become productive participants. At first glance, the lack of access to the work practices of others found in Zooniverse and other citizen science project may appear as an anomaly. However, the same characteristic applies to a greater or lesser degree in most 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 up to a page edit could be a few seconds to correct a typo, or hours of research and fact checking. Similarly, in Free/Libre Open Source projects [4] participants can see the code 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 may benefit from high quality practice proxies to support learning and knowledge sharing.

To improve newcomer learning, designers for open online communities might consider ways in which to effectively insert richer practice proxies into the newcomers' experience to help them orient themselves to 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 support their continued participation in 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 explore how presenting practice proxies to newcomers at different moments relates to their growth in self-efficacy, learning, and motivation. 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. Online community design promoting social translucence might seek to support practice proxies rich in contextual detail and specificity about participants' practice. Simply providing access to the decisions made by a participant may not be sufficient to create social translucence in system design. By having practice proxies that emphasize context and specificity, a system has the potential to embed a richer description of community practice norms, thus helping a newcomer gain a deeper understanding of how to be a full and valuable member. Moving forward, we plan to further explore 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 explored how new users orient themselves towards the goals and practices of massively distributed online collaborative projects. By combining the theories of legitimate peripheral participation [18] and social proxies [8, 9] we focused on the relationship between access to participate and access to others' participation in online environments. Empirically we examined two citizen science projects where the traces of primary practice are not accessible, yet where 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 traces of participation 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 serves as resources for new users as they learn how to become participants in the project.

By combining the concepts of social proxy and LPP we can seek out design approaches that better orient new users towards the goals and practices of online crowdsourced projects. To further this particular design approach to online crowdsourced work, future research should consider design interventions that test the relationship between practice proxies and learning opportunities for new users. Building on rich qualitative and trace data, such experiments might involve testing the degree of access to practice proxies and measuring the impact this has on continued participation.

ACKNOWLEDGMENTS

Without the Zooniverse volunteers working on the projects and taking the time to answer our questions there would be no paper. Many thanks to Stuart Lynn and the Zooniverse team for access to PH and SE data. Several colleagues and the anonymous reviewers offered invaluable feedback. This material is based on work supported by the National Science Foundation under Grant No. IIS 12-11071.

Kevin Crowston is supported by the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

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. *Congition 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.