Please Say "Shibboleth": Socialization Through Language Adoption in Virtual Citizen Science

Corey Jackson

The Information School, The University of Wisconsin- Madison
corey.jackson@wisc.edu

Abstract

Socialization is crucial for open collaboration projects as incumbent participants establish and build relationships within their community. However, socialization often excludes participation in a group's communicative and language practices. This study examines the socialization of newcomers to the communicative and language practices of Gravity Spy. Participants classify image subjects in Gravity Spy to isolate noise signals in gravitational wave detectors. Analyzing the discussion posts of Gravity Spy newcomers, we determine how they assimilate into the group's communicative practices. We find that socialization in the language practices of a group can be a valuable metric of integration. The results suggest successful socialization is marked by the convergence of newcomers' language with the community, influencing retention and participation. These findings provide insights into the linguistic dynamics of virtual citizen science communities and propose strategies to improve the integration of newcomers through language adoption.

Introduction

Open collaboration is a form of interaction conducted over the Internet where people separated by time and distance collaborate to produce artifacts. Forte and Lampe (2013) define open collaboration as "an 'online environment that (a) supports the collective production of an artifact (b) through a technologically mediated collaboration platform (c) that presents a low barrier to entry and exit and (d) supports the emergence of persistent but malleable social structures." Wikipedia and F/LOSS are but a few examples of open collaboration projects. Many open collaboration projects consist of virtual groups separated by time and space, whose interaction is mediated via information and communication technologies. On Wikipedia, people from around the world collaborate on articles, and in F/LOSS projects, people write and debug software code written by other developers. The success of virtual communities depends, in part, on the management of contributors, specifically newcomers who often struggle to enter new groups (Kraut et al. 2012).

The management of newcomers is not unique to open collaboration. The challenges faced by newcomers and those who manage the group run parallel to those faced by managers or employees in organizational settings, which have been addressed through socialization. Socialization is the process by which incumbents acquire attitudes, behavior, and knowledge to assimilate into groups (Maanen, Eastin, and Schein 1977). The literature on organizational socialization has provided valuable evidence for the fields of human-computer interaction (HCI) and computer-supported collaborative work (CSCW) to enhance our understanding of processes and strategies to ease entry into the group. To help assimilate incumbent members, they should undergo training. We also know incumbent members are often proactive by seeking information from more established group members (Bateman, Gray, and Butler 2011). Despite the usefulness of the organizational socialization literature, the form and functioning of open collaboration platforms mean innovative tactics, measurements, and outcomes may be required to evaluate socialization effectively in this context.

We contend that socialization in a group's language practices could be valuable in determining socialization. Language is crucial to how groups coordinate, interrelate, and adapt (Swol and Kane 2019). Language may provide hints about micro-level (i.e., individual speaker) and macro-level behaviors (i.e., group phenomenon that implicates social structures and processes). The acquisition of language and culture is a mutually constitutive process (Ochs and Schieffelin 2008). Coherency in a group's language norms and practices increases the fluidity of interaction among speakers, facilitates cooperation toward shared goals, improves language quality, and promotes conversational grounding (Bjørn and Ngwenyama 2009). Furthermore, language studies have shown that a shared language among a group's members is crucial for identity formation and retention (Schieffelin and Ochs 1986; Ochs and Schieffelin 2008; Bjørn and Ngwenyama 2009; Goldberg and Srivastava 2017).

This paper examines the linguistic contributions and behaviors of Gravity Spy, a virtual citizen science project. Citizen science involves amateurs in various aspects of scientific research, including data collection, data analysis, and writing up the results of analysis projects (Bonney et al. 2009). Virtual citizen science (henceforth VCS) engages amateurs through the Internet. In Gravity Spy, people volunteer their time to contribute to science by reviewing and classifying image subjects produced by instrumentation used to detect gravitational waves. In addition to classifying image subjects, volunteers participate in discussions about the data and auxiliary topics. The conversations on discussion boards are crucial for coordinating work, negotiating the appearance of phenomena in the data, and discussing scientific debates among volunteers (Jackson et al. 2018; Mugar et al. 2014). Furthermore, citizen science projects are increasingly shifting towards tasks requiring higher cognitive complexity and sophisticated communication. Volunteers frequently engage in interpretive and analytical activities, such as synthesizing evidence from multiple datasets to draw scientifically meaningful conclusions or making judgments that inform subsequent data collection strategies. For example, in Zooniverse's Galaxy Zoo project, volunteers not only classify galaxies but also discuss and collaboratively interpret astrophysical phenomena—such as the discovery and analysis of unusual galaxies like "Green Peas"—to guide the scientific community's attention toward new research avenues (Keel et al. 2012; Lintott et al. 2009; Cardamone et al. 2009). Thus, as tasks become more cognitively demanding, effective participation increasingly relies on volunteers' ability to communicate their reasoning clearly to peers and professional scientists. In that sense, building and demonstrating communicative competence is a socialization metric (or shibboleth) for varied aspects of practice in citizen science.

Beyond citizen science, linguistic assimilation is critical across where sustained participation and effective coordination are needed. For example, in open-source software communities like GitHub or GitLab, newcomers must rapidly assimilate technical jargon and interaction norms to contribute code and engage in collaborative debugging meaningfully. Similarly, online learning environments, including MOOCs (Massive Open Online Courses), require participants to adapt to educational discourse conventions and terminologies to foster productive interactions and peer support. Online discussion spaces around specialized topics, such as medical forums (e.g., PatientsLikeMe), advocacy groups (e.g., climate action communities), or hobbyist platforms (e.g., Stack Overflow), also necessitate linguistic adaptation for successful integration.

This research explores linguistic adoption by closely examining linguistic socialization within Gravity Spy, a Zooniverse-based citizen science project. This research provides insights into broader implications across varied forms of open online collaboration.

Language Socialization

Despite extensive research on socialization, less is known about how newcomers assimilate into a group's language practices or develop language socialization behaviors. This study draws on literature describing socialization in open collaboration groups, emphasizing the role of language in coordinating actions, fostering relationships, and enabling group adaptation (Swol and Kane 2019). Establishing norms of communicative behavior enhances community interaction, offering sociability, support, and a sense of identity (Lam 2008). Developing communicative competence is essential for group functionality. On Wikipedia, "meta-workers" coordinate much of their work on Talk pages, yet discussions often pose a barrier for newcomers.

Language socialization involves how novices gain communicative competence, group membership, and legitimacy (Duff 2007, p. 310). Research has explored language use across various virtual groups, such as Usenet (Stewart et al. 2017), fandoms (Burke, Kraut, and Joyce 2009), multiplayer games, and social networks (Ahuja and Galvin 2003), revealing social and technical dynamics in participation. Key areas of focus include membership claims (Burke, Kraut, and Joyce 2010), social accommodation (Giles 1973; Danescu-Niculescu-Mizil and Lee 2011; Moon, Potdar, and Martin 2014), language change and variation (Nguyen and Rosé 2011), and language use's link to continued participation (Danescu-Niculescu-Mizil et al. 2013).

Research shows that people adjust their communicative behaviors to align with group norms and demonstrate affinity. For example, in Wikipedia, low-status editors converge linguistically with high-status editors, while in MOOCs, students mirror the language of student leaders (Moon, Potdar, and Martin 2014). The theory of social accommodation (Giles 1973) explains such adjustments as efforts to build solidarity or differentiate oneself. Language convergence is a strategy for gaining social approval and improving group cohesion and performance (Swol and Kane 2019).

Language use in virtual communities evolves as the community develops. Danescu-Niculescu-Mizil et al. (2013) found a two-stage linguistic lifecycle in online beer rating communities: an initial phase of innovation followed by linguistic conservatism. Similarly, Nguyen and Rosé (2011) observed that language stabilizes after 8–9 months of participation. Studies have also identified persistent shifts in language use over time. For instance, Danescu-Niculescu-Mizil et al. (2013) noted that high-status Wikipedia editors became less polite after gaining status. Ahuja and Galvin (2003) highlighted differing communication dynamics between newcomers and experts, with experienced members often developing forum-specific jargon and informal interaction styles (Nguyen and Rosé 2011).

Language socialization and behavior Research links language practices to behavioral outcomes, including discourse quality (Matthews et al. 2015), member satisfaction (Arguello et al. 2006), retention (Niederhoffer and Pennebaker 2002), and task success (Cassell and Tversky 2005). Language socialization provides insights into group assimilation, showing that early adopters of community language have longer tenures (Danescu-Niculescu-Mizil et al. 2013). Posts that are topically relevant, ask questions, or use simpler language garner more replies, which promote continued participation (Duff and May 2017). Email studies reveal reduced language use among members planning to leave (Goldberg and Srivastava 2017). Similarly, participants who increasingly use "we" and reduce "I" demonstrate growing community identification (Burke, Kraut, and Joyce 2009). In Twitch chats, non-subscribers use more emotional language, while subscribers favor analytical expressions. Increased use of first-person plural ("we") and positive emotional words correlates with higher member satisfaction (Matthews et al. 2015), making language behaviors predictive of user engagement.

In citizen science, language socialization is unique due to several factors. Projects often lack formal authority figures to codify linguistic practices, and participants typically lack scientific training, making jargon less accessible. Communication is loosely regulated, with ever-growing, varied user-generated content. Additionally, asynchronous participation allows newcomers to remain invisible, bypassing traditional discursive spaces. The "reader-to-leader" framework describes this transition, where participants initially lurk before gradually contributing and becoming central to the community (Luczak-Roesch et al. 2014). These dynamics make citizen science a compelling context for studying language socialization.

Language in digital citizen science Research on language in volunteer science is limited. Studies on Zooniverse projects have shown that language evolves even within the same scientific domain, shaped by technical task challenges and the establishment of shared scientific vocabulary in early project stages (Luczak-Roesch et al. 2014; Rohden et al. 2019). Volunteers navigate existing norms while shaping language practices, highlighting the need to understand mechanisms of newcomer socialization in this context.

Citizen scientists must learn community-specific language to engage in discussions and perform advanced tasks like data analysis and paper writing. Linguistic competence enhances conversational grounding and task execution (Bjørn and Ngwenyama 2009). Discussion pages in citizen science projects serve diverse functions, including posing questions, providing feedback, and soliciting information. While some projects maintain stable language (e.g., astronomy-focused projects (Luczak-Roesch et al. 2014)), others develop complex jargon as practices evolve. This dynamic can challenge newcomers, who may struggle to understand and adopt linguistic norms. Volunteer-led discoveries like Hanny's Voorwerp (Lintott et al. 2009; Keel et al. 2012) and green peas (Cardamone et al. 2009) exemplify how evolving language supports increasingly complex tasks.

Newcomers influence and are influenced by a group's language practices, shaping language dynamics in linguistically innovative projects (Jackson et al. 2019b). Unlike structured training for classification tasks, no formal tutorials exist for learning a project's linguistic norms, leaving newcomers to navigate language acquisition independently. Language socialization parallels the assimilation processes seen in children and Usenet communities, where shared language fosters group identity, conversational grounding, and retention (Schieffelin and Ochs 1986; Ochs and Schieffelin 2008; Bjørn and Ngwenyama 2009).

This raises key questions addressed in this manuscript **To what degree do newcomers innovate versus mimic a community's linguistic practices?** and **How does linguistic composition affect volunteers' willingness to continue participating?**

Setting

This research focuses on newcomers to Gravity Spy, a VCS project hosted on the Zooniverse platform (Simpson, Page, and Roure 2014). A detailed description of Gravity Spy can be found in (Zevin et al. 2016). Gravity Spy participants help astrophysics search for gravitational waves and identify and isolate noise signals generated during observation runs. The classification interface is depicted in Figure 1. To submit a classification, participants judge which, if any, reference glitches match the image subject. Since new instrumental and environmental noise signals can emerge, sometimes no reference glitch reaches the current subject; in such cases, participants are encouraged to select the "None of the Above" option. Each image subject received several rounds of human judgment, and the most oft-selected reference glitch class for an image subject is submitted as the answer.

Research on Gravity Spy and other Zooniverse projects highlights the importance of discussion boards for participant engagement (Jackson et al. 2019a, 2018; Luczak-Roesch et al. 2014; Jackson et al. 2019b; Rohden et al. 2019; Mugar et al. 2014). These boards facilitate behaviors such as sharing work practices (Mugar et al. 2014) and producing knowledge (Jackson et al. 2018). For example, discussions on Planet Hunters serve as a learning tool for newcomers by offering information on classification practices (Mugar et al. 2014). Similarly, participants in Gravity Spy use discussion boards to identify and curate new glitch classes, which require them to describe morphological features in ways understandable to both peers and professional scientists (Jackson et al. 2018). These discussions often involve creating new vocabulary and adding an additional layer of specialized terminology that newcomers must learn beyond basic scientific jargon.

Communicative and language norms on discussion boards are often unclear to newcomers. When Gravity Spy launched, few resources explained how to participate in discussions. As the content grows, participants introduce new linguistic styles, making it harder for newcomers to assimilate. This lack of guidance highlights the need for studies that examine language and communicative patterns in these projects to better understand how newcomers navigate and contribute to these critical interactions.

Methods

Our primary units of analysis include community and user language practices aggregated by week.

Data Collection

We collected data from the Talk forums in Gravity Spy. Talk forums are the primary medium for participants to communicate with each other, project scientists, and software developers. Each comment record contained the user's ID who made the comment, the comment body, and other metadata, such as timestamps indicating the exact date the comment was posted to the system. We also collected data to determine when a participant first joined the project, as some volunteers may participate in classification before posting com-

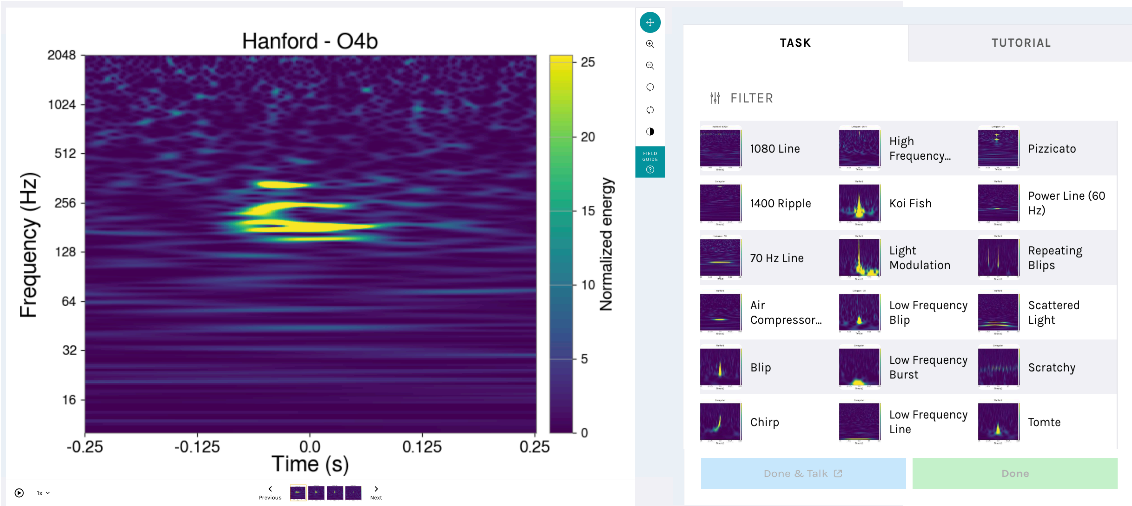

Figure 1: The classification interface. Volunteers are presented with a spectrogram and asked to categorize the noise signal represented in the image.

ments. The data set contained 2,833 users who contributed to Gravity Spy since March 3rd, 2016. The last record was posted on January 21st, 2024. We transformed the data to ensure the ecological validity of our analysis, removing the final week since the collection ended before the last day of the previous week in our data set.

Our analyses intentionally excluded measurements from project organizers and beta testers to ensure that our findings accurately reflect the linguistic assimilation processes characteristic of general volunteers. Project organizers typically possess specialized expertise and distinct linguistic patterns that differ substantially from general participants, which could artificially inflate linguistic assimilation measures and distort community dynamics analyses. We also removed data from users who joined and contributed during the project's testing phase, March 3rd, 2016 and October 12th. After removing these data, the data of 3,076 users contributing over seven years (or 379 weeks) remained.

We chose week as a temporal window to capture changes in language practices because days and sessions can introduce too much variability and noise, making it difficult to discern meaningful patterns. These data may be overly granular, leading to potential overfitting in our models due to short-term fluctuations in user behavior. In contrast, months may smooth out important fluctuations and trends, potentially overlooking significant variations and interactions that occur on a shorter time scale. By aggregating user data by week, we struck a balance between capturing sufficient detail to understand user engagement and linguistic assimilation while minimizing the noise and variability inherent in daily observations.

High-Level Language Categories

The challenge becomes to distill the language. Our primary analysis method examines the interactional and discursive patterns exhibited by the community at large and individual volunteers. This requires analysis of comments and parsing through the content to identify key themes and trends. Text cleaning procedures removed noise and standardized the language used in comments. This included removing stop words, normalizing punctuation and capitalization, and stemming or lemmatizing words to their root forms. Our analysis extracted important features for understanding user behavior and linguistic patterns. These features can be grouped into three high-level categories: temporal features, content features, and interaction features, described below.

Temporal Features: capture the aspects of time and participation duration, reflecting how users' engagement evolves. These predictors include:

- `week` - the week the user contributed.
- `weeks since joined` - the number of weeks since the user joined the project.
- `days` - the number of days the user contributed in a given week.
- `participation gap` - the number of weeks between the current and previous week commenting, accounting for non-consecutive participation.

Content Features: describe the nature and type of user contributions, indicating the diversity and richness of the content they produce. These predictors include:

- `questions` - the number of questions in a post as indicated by the appearance of "?" punctuation mark[1].
- `links` - the number of URLs in a post as indicated by the appearance of the "http" or "https" schema.
- `tags` - the number of tags used by the user as indicated by the hash (#) sign

[1]We exclude instances where "?" appears in URLs

- `user references` - the number of times the user references other users indicated by the presence of the @ symbol

Interaction Features: focus on the user's engagement with the community and the social dynamics of their participation. We parsed comment data to extract hashtags by matching a regular expression for:

- `comments` - the number of comments the user posts.
- `threads started` - the number of new discussion threads started by the user.
- `token innovations` - the number of new words the user introduces.

Measuring language assimilation Language assimilation was measured using cosine similarity. Cosine similarity is a measure of similarity between two sequences of numbers and is measured by the cosine of the angle between two vectors. In the equation below, documents A and B are described by vectors that include the frequency term of n words, where A is the volunteer's words, B represents the language dictionaries, and n is the total number of individual words in the corpus. The cosine values range from -1 to 1, where a score of 1 is perfectly similar to the language of the community, and a score of -1 is perfectly dissimilar.

We computed the cosine similarity scores for each volunteer for each week in the project. To compute cosine similarity, we first tokenized each user's comments and created a vector (A) representing the presence or absence of each token in the user's language. The community language was derived from a process that considered only popular tokens (`viral tokens`) during that week and removed tokens that had fallen out of the normal conversation of the project. We developed a naive approach to identify non-popular tokens (`nonviral tokens`) by considering tokens that did not appear at least once by another user in the previous thirty days and by at least five users in the week being evaluated. Including `viral tokens` ensures that users are not penalized for not learning outdated language.

In computing the community language, we excluded each user's contributions to avoid artificially inflating the representation of the community language. The resulting `viral tokens` formed vector B. We then matched vector B with vector A by filling vector A with 1s and 0s, where 1 indicates the occurrence of the word in the user's comments, and 0 indicates the absence of the word. This process created two vectors of equal length for the calculation of cosine similarity. Using these criteria, our algorithm iterated over the data for each user in each week they contributed to the project. This allowed us to accurately measure the alignment of each user's language with the community language every week. This produced three additional interaction features.

- `viral tokens` - the number of tokens used to compute similarity.
- `nonviral tokens` - the number of tokens not deemed viral.
- `cosine similarity score` - the similarity between the user's language and the community language (or viral tokens).

Results

Before addressing our research questions, we wanted to understand the characteristics of the linguistic phenomenon at the community level. As noted in Section , users make significant linguistic contributions and innovations in Gravity Spy. During the period represented in our data, 2,833 users made 130,020 comments.

Gravity Spy's Linguistic Dynamics

To estimate the linguistic health of Gravity Spy, we evaluated the growth in different units of linguistic interaction measures. The scatter plots on the left side of Figure 2 depict the weekly production of new comments, threads, tokens, and users. This smoothed line helps to visualize the overall trajectory and changes in linguistic activity, smoothing out short-term fluctuations to highlight long-term trends. However, the three bottom scatter plots show similar data in the aggregate. The left charts depict the cumulative growth of these same features, again showing the project launch week and a smoothed line.

Visual inspection of the graphs and the smoothed line reveals interesting trends related to the project's linguistic dynamics and user engagement patterns. First, the project launch week (12 October 2016) and approximately two years later (October 2018) are marked by high contribution volumes. This appears to be where most of the project activity occurs, indicating a period of intense engagement and linguistic expansion. Second, around April 2020, the project experienced significant increases in the weekly production of comments, new threads, and new tokens. Additionally, the number of new users joining Gravity Spy lasted for several weeks in April 2020. We also observe two additional periods of notable increases in new tokens around April 2021 and October 2022, indicating bursts of linguistic innovation and possibly introducing new topics or activities within the project. Third, the volume of tokens contributed after the April 2020 period appears to increase and remain steady compared to before April 2020.

Some trends (though not all) are unsurprising and can be attributed to events surrounding the broader scientific community or the project. For example, significant increases in contributions often coincided with major discoveries, publications, or media coverage related to the project's scientific goals. For example, the increases in contributions and users during x coincided with the discovery of gravitational waves (GW150914, GW170817, GW170814, and GW170104) by the LIGO community and the media coverage of the discovery. Additionally, when LIGO conducts new observation runs, new images are uploaded to Gravity Spy. These external factors likely spurred renewed interest and participation among volunteers, leading to bursts of activity and linguistic contributions.

We converted the data to a time series to understand the temporal evolution of project dynamics, removing data from the testing phase and project organizers. We used Pettitt's Test for Change-Point Detection and the Theil-Sen median slope estimator to statistically evaluate Gravity Spy's activity patterns and detect significant shifts over time. The results are reported in Table 1.

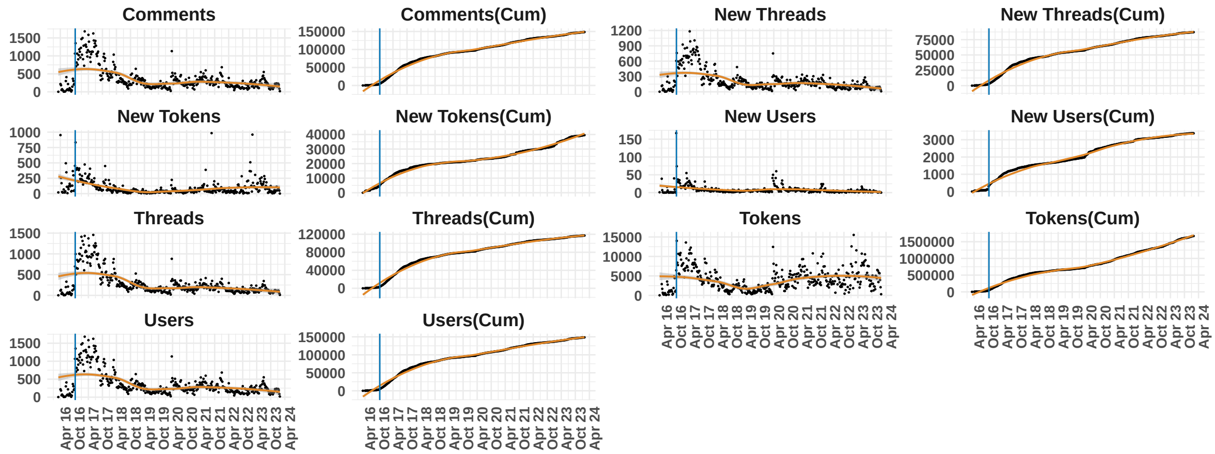

Figure 2: Weekly activity metrics of the Gravity Spy project from October 10, 2016, to January 2024. The plots illustrate the weekly additions and cumulative counts (Cum) of new comments, threads, tokens, and users. The blue vertical line represents the official launch date of the project.

Compared to the start of the project, most measures showed early changes and significant downward shifts soon after estimates from the project launch week (10 October 2016). This is unsurprising given the visual inspection of the scatterplots in Figure 2. New comments and threads peaked during the week of February 12th, 2017 (17 weeks after the project launched) with a Sen's Slope Estimator of -1.20 and -0.70, respectively. Perhaps unsurprisingly, the volume of new tokens in the project tended to show an earlier change occurring ten weeks after the project launched (December 25th, 2016); however, the decreasing rate of change was much lower than other measures (-0.07). Perhaps most important for sustaining community growth, new users joining only began to shift around on June 11th, 2017, or 34 weeks after the project launched and having a relatively small change rate (-0.02). Only aggregate tokens had a positive trend (0.4).

	Change-Point		Sen's Slope	
Measure	*Statistic*	*Estimate*	*Statistic*	*Estimate*
New Comments	24704	128 (2017-02-12)	-12.21	-1.20
New Threads	22386	131 (2017-02-12)	-12.64	-0.70
Threads	27121	131 (2017-02-1)	-14.65	-1.06
New Tokens	16451	78 (2016-12-25)	-2.83	-0.07
Tokens	14207	78 (2016-12-25)	0.27	0.4
New Users	18627	247 (2017-06-11)	-9.92	-0.02
Users	24704	128 (2017-02-12)	-12.21	-1.2

Table 1: The table above displays the results from the Pettitt change-point and Theil-Sens estimator for community linguistic measures in Gravity Spy. The statistics were calculated beginning on October 10, 2016, the week Gravity Spy was launched.

The overview of Gravity Spy's linguistic health reveals a dynamic and evolving language landscape with an increasing cadre of users. For newcomers, the growth means they must quickly learn a substantial amount of new terminology and adapt to the community's linguistic norms. Indeed, users who joined during the growth period may have to contend with a rapidly expanding vocabulary and evolving discussion topics, while users joining after the growth period may have to learn fewer new terms but must contend with discussions spread across time and virtual space, which may cause challenges in assimilation into the project's linguistic norms and practices. For example, a volunteer joining on January 1st, 2023, had to familiarize themselves with a specific set of active discussions and terminology, while a user joining on January 1st, 2024, faced a different set of terminology and possibly less active but more dispersed discussions. More established users are concerned about whether they can make sense of the evolving discourse and maintain their engagement and contributions.

User Engagement in Project Discussions

While the overall volume of user contributions to discussions is significant, prior research on user engagement patterns in citizen science has demonstrated that contribution remains unequal - most users are one-time or infrequent contributors (Ponciano and Brasileiro 2014) and even fewer participate in discussions (Nov, Arazy, and Anderson 2011). Of the Gravity Spy population, 2,833 (10.1%) users made comments [2]. Throughout their lifetime, users contribute on average 35 (σ = 522) comments; however, the median is three. Another aspect of participating in discussions is when, from one perspective, volunteers might observe before contributing to practice, and others may choose to jump right in and immediately start contributing. On average, users waited 5 (σ = 21) weeks after they joined the project before com-

[2]We removed project organizers and beta stage testers

menting. However, approximately half of the volunteers who posted comments did so in the same week they joined the project, indicating a mix of immediate and delayed engagement strategies among participants.

Another aspect of participation is the duration of involvement. After initial contributions, the average participant remained active in discussions for 2.5 weeks (σ = 7.21, x = 1). This suggests that while some users may contribute consistently over an extended period, many have brief engagement periods. Understanding these participation patterns can help design interventions to encourage sustained engagement and more balanced participant contribution.

Assimilation Into Community Linguistic Practice

Depicted in Figure 2 are trend lines for contribution volume of measures such as comments and threads and linguistic characteristics such as question-asking, hyperlinking, and tagging conversations and image subjects. Visual inspection of the trend lines reveals that volunteers' average contributions steadily increased during the ten weeks for most measures. We computed Pettitt's Test for Change-Point Detection and Sen's Estimator to know the statistical parameters of the data.

Newcomer linguistic assimilation We extracted the first week's contributions for users having made comments, removing project organizers and testers (resulting in a dataset of 3,076 users). The average volunteer makes 6.17 (x = 2, σ = 15.63) comments, initiates 3.72 (x = 1, σ = 9.91) threads, and has 0.99 (x = 0, σ = 3.25) token innovations. Newcomers were active contributors to classifying with an average of 365 (x = 204, σ = 507) classifications. We evaluated assimilation using cosine similarity. The average similarity for newcomers was 0.08 (x = 0.06, σ = 0.06). Compared to the dataset's average similarity, the newcomer similarity score (0.08) is 3.44 times lower than the average of 0.31 computed weekly.

Next, we examined how temporality impacts newcomers' language assimilation. Controlling for virality, we computed the average number of tokens that comprised the vector to calculate similarity scores for each week. This indicates what words are popular in the project; thus, a newcomer should be expected to learn. We also computed the number of non-viral words (users are exposed to both). During the project's first week, there were 176 viral tokens and 5,457 non-viral tokens; one year later, the number was 1,537 and 21,116, respectively. This means that after one year, there were 8.73 times more viral words and 3.87 times more non-viral words. In the final week of observation (2024-01-21), there were 3,692 viral tokens and 80,604 non-viral tokens; 20.98 times more viral words than when the project started and 14.77 times more non-viral words. This indicates a substantial increase in both viral and non-viral language over time, suggesting that the linguistic environment of the project becomes increasingly complex as it evolves.

We fitted a linear regression model to predict cosine similarity with three temporal features - the week the user contributed (week), the number of weeks since the user joined (weeks since joined), and the number of days the user contributed that week (days). The model explained a significant portion of the variance in cosine similarity scores, R^2= 0.35, F(3, 3072) = 540.3, $p < 0.001$. The model is reported in Table 2, M1. The model reveals that time significantly influences linguistic assimilation. The coefficient for the week variable (β = -0.00003, $p < 0.001$) suggests a small but statistically significant negative relationship between the time since the project started and the cosine similarity score[3]. This implies that users' language becomes slightly less similar to the community language as the project progresses. The number of days contributed also impacts the similarity score. The coefficient for the number of days contributed per week (β = 0.03, $p < 0.001$) was positive and statistically significant. This suggests that users who contribute more frequently within a week tend to have higher cosine similarity scores. The period between joining and making an initial contribution did not significantly impact similarity (β = -0.00005, p = 0.29).

Our central hypothesis is that during the early stages of a user's participation in the project, formative experiences might predict socialization into linguistic norms and the community. We examined newcomer retention as a function of their linguistic and activity behaviors during their first week contributing. Of the 3,076 users, 2077 (67%) dropped out after the first week. A t-test revealed a significant difference in cosine similarity scores between users who dropped out (0) and those who returned (1) after their initial week contributing, t = -12.43, df = 1400.4, $p < 0.001$). The mean cosine score for dropout users was 0.07, while the mean for retained users was 0.10.

Next, we fitted a logistic regression to determine how the linguistic and activity measures predicted retention (M2, Table 2). In addition to temporal predictors, M2 includes content-related predictors (e.g., questions) and interaction-related predictors features (e.g., threads started). The model performed well, with the likelihood ratio test comparing the fit of the model against a null model being significant ($\chi^2(14)$ = 270.58, $p < 0.001$), indicating that the model with predictors provides a significantly better fit to the data than the null model with no predictors.

The model reveals that newcomers who waited to post a comment were ultimately more likely to return to the project (β = 0.0005, p = 0.01) and that the number of days they contributed during their first week posting comments had a significant impact on the likelihood of returning to the project (β = 0.052, $p < 0.001$). Surprisingly, only one activity variable was shown to significantly influence the likelihood of retention, suggesting that posting links positively influenced retention (β = 0.046, p = 0.03). Finally, the coefficient for cosine similarity (β = 5.20, $p < 0.001$) was positive and significant, indicating that users whose language closely aligns with the community language during their first week contributing to comments are more likely to be retained.

Overall, these results suggest that as newcomers join the

[3]Since the dependent variable (cosine similarity) is bounded in such a narrow range, small changes in the predictors can lead to significant changes in the outcome. Therefore, small coefficients can still have meaningful effects on the outcome variable.

	Dependent variable:		
	Newcomer Cosine	Newcomer Retention	Cosine
	OLS	*logistic*	*linear mixed-effects*
	(**M1**)	(**M2**)	(**M3**)
Temporal Features			
`Week`	-0.0000(0.0000)***	-0.0003(0.001)	-0.04(0.01)***
`Weeks Since Joined`	-0.0000(0.0000)	0.01(0.002)*	0.03(0.002)***
`Days`	0.03(0.001)***	0.52(0.07)***	0.01(0.001)***
`Participation Gap`			-0.005(0.001)***
Content Features			
`Questions`		-0.02(0.02)	0.002(0.002)
`Links`		0.46(0.21)*	0.002(0.001)*
`Tags`		-0.01(0.01)	-0.01(0.002)***
`User References`		0.25(0.15)	0.002(0.001)**
Interaction Features			
`Comments`		0.03(0.02)	0.02(0.003)***
`Threads Started`		-0.02(0.02)	
`Token Innovations`		0.01(0.02)	0.0000(0.001)
`Viral Tokens`		-0.001(0.0004)	-0.05(0.004)***
`Non-Viral Tokens`		-0.0000(0.0000)	0.06(0.004)***
`Cosine Similarity`		5.19(1.32)***	
Constant	0.62(0.02)***	2.14(8.78)	0.14(0.002)***
Observations	3,076	2,154	8,648
R^2	0.35		
Adjusted R^2	0.34		
Log Likelihood		-1,222.94	10,289.20
Akaike Inf. Crit.		2,475.89	-20,542.40
Bayesian Inf. Crit.			-20,415.23
Residual Std. Error	0.05 (df = 3072)		
F Statistic	540.35*** (df = 3; 3072)		
Note:			*p<0.05; **p<0.01; ***p<0.001

Table 2: Models for predicting newcomer cosine similarity and retention in the Gravity Spy project. Model 1 (M1) uses Ordinary Least Squares (OLS) to predict newcomer cosine similarity, Model 2 (M2) employs logistic regression to predict newcomer retention, and Model 3 (M3) utilizes a linear mixed-effects model to predict cosine similarity.

project, they face a growing and dynamic vocabulary that they need to assimilate. The increase in both viral words implies that newcomers must not only learn the language of the community but must also keep up with (or sift through) the evolving trends and niche terms that become part of the project's discourse.

Long-term assimilation Next, we combine more of the data to paint a more robust picture of how linguistic measures evolve and their influence on the behavior of users. This analysis uses the user/week dataset, meaning users can have multiple records. This dataset contains 8,831 records; 1,595 were removed because they were project organizers or testers. In addition to the features described in the previous section, we computed the participation gap, the number of weeks between the current and prior week's commenting, to account for the fact that users do not contribute every week, and the community may have shifted. We also included user week, a variable sequenced to the occasions users contribute to discussion boards.

Few users sustain contributions over an extended period, with only 136 users having ten weeks of posting (these can be non-consecutive); only 13 users have contributions in 100 weeks. We used features similar to those in the previous analysis to predict the cosine score. Since the data are repeated measures, we fitted a linear mixed-effects model to account for the within-user correlation by including a random intercept for each user. This approach allows us to model the variability both within and between users. The likelihood ratio test compares the model's fit against a null model with no predictors. The test was significant (Cox and Snell (ML) $R^2 = 0.25$, $\chi^2(15) = 2444.1$, $p < 0.001$), indicating that the model explains approximately 25% of the variance in the outcome variable and the model with predictors provides a significantly better fit to the data than the null model.

The results in Table 2, M3 reveal that as the project progresses, users' weekly cosine similarity scores decrease slightly (β = -0.0422, $p < 0.001$), the longer a user has been part of the project, the higher their weekly cosine similarity scores (β = 0.03, $p < 0.001$), the more frequent participation within a week is associated with higher cosine scores (β = 0.01, $p < 0.001$), and gaps in participation are associated with lower cosine scores (β = -0.05, p $<$ 0.001). We also find that the more a user comments (β = 0.0239, p $<$ 0.001), posts links (β = 0.02, p = 0.042), and refers to other users (β = 0.02, p = 0.009), the higher their cosine scores. The appearance of tags (β = -0.01, p $<$ 0.001) and classifications (β = -0.01, p $<$ 0.001) was associated with lower cosine scores. We also examined how the existing corpus of tokens influenced users' linguistic assimilation. The model shows that more viral tokens lead to lower cosine scores (β = -0.0524, p $<$ 0.001).

Discussion

Our analysis reveals key aspects of community and individual user language dynamics in Gravity Spy. At the community level, linguistic trends show significant growth during the first two years, followed by a decline influenced by external factors, such as LIGO discoveries. These events spurred bursts of linguistic activity, resulting in an increasingly complex vocabulary. This complexity poses challenges for newcomers' assimilation while requiring experienced users to continuously adapt to maintain effective communication.

The growing complexity of Gravity Spy's linguistic ecosystem makes active and frequent participation essential for newcomers to integrate. Language convergence, often driven by social approval or affiliation (Giles, Coupland, and Coupland 1991), plays a vital role in supporting group cohesion and retention.

At the individual level, linguistic assimilation is more influenced by temporal factors than community dynamics. Frequent contributions (`Days`) improve assimilation by allowing users to observe and adopt language norms, while time since project launch (`Week`) decreases assimilation due to increasing language complexity. Immediate engagement after joining appears less critical than sustained participation, as one-time users often lack the background to assimilate effectively. Over time, the growing complexity of community language further hinders newcomers, raising concerns about retaining engaged participants as the project evolves.

Longer-term assimilation for users who contribute more frequently, as well as many temporal, content, and interaction measures, is crucial for predicting continued language assimilation (M3). Temporal measures such as `Week` and `Participation Gap` contribute to decreasing language assimilation. The former is likely related to the increasing complexity and volume of community language over time, making it harder for newcomers to keep pace. The `Participation Gap` means that the community language has evolved during the user's absence, potentially leaving experienced users behind after taking a break from the project. Interestingly, the question of lurking (measured through `Weeks Since Joined`), while not a factor for newcomers in M1, has a positive influence on language assimilation. Again, this may be explained by the fact that lurking allows users to observe and learn the community norms and language before actively participating. Unsurprisingly, `Days` contributing during the week also contributes to greater assimilation.

The content of users' comments provides insights into their assimilation process and engagement patterns. In Gravity Spy, prior research has noted shifting practices in the discussion threads where more experienced users become responsible for linguistic innovations as they engage in activities such as anomaly searching, curating, and classifying new noise signals. The results of M3 support this practice shift as tagging (#) is associated with decreased assimilation. In the case of Gravity Spy, Jackson et al. (2018) described how personomies (or the language unique to a volunteer's practice) are often hidden but, over time, might become visible as folksonomies. This practice shift contrasts the more static language used in less dynamic projects. More collaborative activities in the project appear to support assimilation - users having a larger volume of `Links` and `User References` in their conversations demonstrate more significant language assimilation.

Concerning newcomer retention, while the number of days contributed and posting links are important, the key determinant of retention is how well a newcomer's language aligns with the community's linguistic norms and practices (M2). Language convergence represents a mode of interaction where newcomers seek to gain social approval by mimicking the communication style of experienced participants. Language convergence has increased group cohesion and performance (Swol and Kane 2019). The `Cosine Similarity` estimate was noteworthy, having the most significant influence on newcomer retention in the model. This suggests that efforts to support newcomers in adopting the community's language could be crucial for improving retention rates.

Building Communicative Competence

Given the diverse backgrounds and varying times at which participants join, how can we ensure newcomers are included in communicative practice irrespective of when they enter? The system, specific project, and science goals in Gravity Spy mean that newcomers face unique challenges in language assimilation. For more experienced users, given the shift in project dynamics and the evolving nature of discussions, supporting different modes of participation in project discussion spaces may be necessary. These are common challenges for many open collaboration platforms and point to the need for strategies and tools to help newcomers and experienced users assimilate and integrate into the linguistic practices of the community.

We propose several strategies and uses of artificial intelligence (AI) focused on standardizing and intelligently enhancing communicative norms and practices to make them less opaque. First, tutorials in citizen science projects are often created to instruct newcomers in completing the project's primary task. In Gravity Spy, users learn to classify with instructional clips on selecting morphology, using tooltips, and best practices for preserving data quality. Tutorials focused on discussion boards might help guide newcomers through terminology and phrases related to the science of gravitational wave physics and emergent project language (i.e., language beyond the subject matter). As the results noted, language changes so that these tutorials might lever-

6. Possible scattered light related glitches:
   6.1. gearwheel
   #gearwheel
   scattered light reminding arches or curves with beaded pattern, accompanied by low frequency burst sometimes.
   6.2. rainbow
   #rainbow
   strong parallels in mid frequency range, high arches, sometimes long and sinusoidal.

7. fuzzy arches
   #wildfire #scratchybridge #wheat
   weak arches in low frequency range built up more or less organized small scratches accompanied by low frequency bursts sometimes.
   (8. Eiffel #eiffel Church #church are rare yet to tell in advance if they should be added later)

In the priority I have considered the following aspects:

- Novelty (in degree of difference from existing classes) i.e. new significantly different glitches from already existing classes.
- prevalence in the new data vs prevalence in old data (O1 vs ER10/O2), those which are present in both are the most important
- prevalence of the ambiguous glitches in the new data which are difficult to classify because they are similar to existing classes but not enough to classify them like that.

Every comment on this is very welcomed.

Figure 3: Example of a discussion page on the Gravity Spy project. The page lists potential scattered light-related glitches, such as gearwheel and rainbow glitches and fuzzy arches, along with their descriptions and tags. The discussion also outlines the criteria for prioritizing these glitches.

age the expert knowledge of experienced users with intimate knowledge of the community's language norms. Projects might also leverage the artifacts that experienced users create over time. For example, users often post about classifying using communicative structures. Mugar et al. (2014) describes how comments made by volunteers in Planet Hunters frequently contain elements of context (what factors contributed to classification) and specificity (where in the image a phenomenon can be located), which the authors refer to as "practice proxies." The language of practice proxies might provide helpful scaffolding for how users should write about the data.

Additionally, one finds artifacts created by experienced users to help convey the structure of the community's evolving linguistic landscape. In an example from Gravity Spy, one user attempts hierarchical categorization of hashtags (Figure 3). These posts refer to specific names of glitches that newcomers might learn.

Building upon this idea, when creating tutorials and other auxiliary language learning artifacts, developers might leverage artificial intelligence to make these materials dynamic, adapting materials to temporal changes in language norms and practices. For example, projects leverage the metadata and trend-based modeling techniques (similar to those used in this research) to determine how language evolves or identifies emergent language (e.g., association rule mining (Ling and Jackson 2024)), which could serve as the basis for scientific discovery. This glossary could include definitions, context of use, and examples.

AI might help by providing contextual suggestions and corrections as users type their comments within the commenting interface. These tools can suggest alternative phrases or correct terminology. Additionally, AI chatbots might assist newcomers by answering questions about terminology and providing examples of proper language use. These are also beneficial in that they can offer real-time feedback on newcomers' posts. While context is often challenging for AI to interpret accurately, existing comments and artifacts can train machine learning models to understand better and provide relevant suggestions. For example, tag gardening has been proposed to maintain the accuracy and relevance of tags within a system (Peters and Weller 2008). This might also leverage existing user-generated artifacts created by users involved in the "infrastructure" and administrative work of projects. For example, one user created a schematic of the relationships among different tags used in the Gravity Spy project (Figure 5). These tools can help enhance the quality of information retrieval and content organization, ultimately improving the ability of users to assimilate.

For some projects, complete language assimilation may not be a goal. In Gravity Spy, experienced users engage in language innovation through their work, creating and proposing new glitch classes. In many citizen science projects, the discussion boards are the primary vehicle for work beyond data analysis (Willett et al. 2013). In Galaxy Zoo, users coordinated much of their work discovering new astronomical phenomena on the discussion boards, creating more discussion and new terminology, such as "green peas." In the case of Gravity Spy, users use the discussion boards to tag glitches, hold discussions, and propose new classifications. This results in new knowledge and language creation, a process described in Jackson et al. (2019a). This new language initially may have few adopters; however, over time, many more users may begin to learn the terminology, describing the personomy to folksonomy shift. In that respect, experienced users who innovate are not causing harm to the community but contribute to the organic evolution of the community language. Models to capture personomies and the transition to folksonomies might be implemented to address this dynamic, ensuring that the evolving terminology is documented and understood within the community context.

The linguistic assimilation dynamics observed in Gravity Spy have implications beyond the immediate context of a single virtual citizen science project. The specific mechanisms identified—such as the significance of linguistic convergence on participant retention—are likely relevant to similar citizen science discussion forums, particularly those hosted on the Zooniverse platform, which shares common communication structures, volunteer dynamics, and participation patterns.

More broadly, these findings potentially apply to other open collaboration communities, including online learning forums, open-source software communities, and domain-specific forums such as health or environmental advocacy groups. However, generalizability should be qualified, as each community possesses unique elements related to domain expertise, task complexity, and norms governing interactions. Future research might empirically test these findings' generalizability by applying similar linguistic analysis methods across multiple Zooniverse projects, such as Galaxy Zoo or Planet Hunters, or by comparing linguistic socialization dynamics between scientific and non-scientific online communities. Additionally, comparative studies between synchronous and asynchronous online forums could further elucidate the boundaries and conditions under which linguistic assimilation significantly impacts participant integration and retention.

Furthermore, while this paper provides insights into linguistic socialization through textual analysis, future research might enrich these findings by incorporating non-verbal communicative cues, such as emoji, images, and other visual artifacts present in online discussions. Additionally, extending the current token-based similarity analysis to include more nuanced linguistic features, such as syntactic complexity, could offer deeper insights into the quality and nature of linguistic assimilation. Finally, insights might be enriched through qualitative data collection, enabling a more comprehensive understanding of why and how participants adopt community language norms and revealing the nuanced social dynamics that purely quantitative measures might overlook.

Conclusion

The principles of language socialization observed in Gravity Spy are relevant to various open collaboration projects and virtual communities. In citizen science projects across the Zooniverse platform, the linguistic assimilation process, where newcomers' language aligns with community norms, is crucial for newcomer integration and retention. Similar dynamics exist in open-source software communities on platforms like GitHub, where new contributors must understand technical jargon and collaborative practices. These findings also apply to online learning platforms like MOOCs, where students engage in forums with subject-specific language, and professional networks like LinkedIn groups, where domain-specific terminology is prevalent. Policy advocacy groups and forums like Reddit or Discord also benefit from improved language assimilation mechanisms, empowering newcomers to contribute more effectively.

Acknowledgments

Without the Zooniverse volunteers who worked on the projects, there would be no article—many thanks to the Zooniverse team for access to data and collaborators who provided valuable feedback on this work.

References

Ahuja, M. K.; and Galvin, J. E. 2003. Socialization in Virtual Groups. *Journal of Management*, 29(2): 161–185.

Arguello, J.; Butler, B. S.; Joyce, E.; Kraut, R.; Ling, K. S.; Rosé, C.; and Wang, X. 2006. Talk to me: foundations for successful individual-group interactions in online communities. 959–968.

Bateman, P. J.; Gray, P. H.; and Butler, B. S. 2011. Research note—the impact of community commitment on participation in online communities. *Information systems research*, 22(4): 841–854.

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

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

Burke, M.; Kraut, R. E.; and Joyce, E. 2009. Membership Claims and Requests: Conversation-Level Newcomer Socialization Strategies in Online Groups:. *Small Group Research*, 41(1): 4 – 40.

Burke, M.; Kraut, R. E.; and Joyce, E. 2010. Membership Claims and Requests: Conversation-Level Newcomer Socialization Strategies in Online Groups. *Small Group Research*, 41(1): 4–40.

Cardamone, C.; Schawinski, K.; Sarzi, M.; Bamford, S. P.; Bennert, N.; Urry, C. M.; Lintott, C.; Keel, W. C.; Parejko, J.; Nichol, R. C.; Thomas, D.; Andreescu, D.; Murray, P.; Raddick, J.; Slosar, A.; Szalay, A.; and Vandenberg, J. 2009. Galaxy Zoo Green Peas: discovery of a class of compact extremely star-forming galaxies. *Monthly Notices of the Royal Astronomical Society*, 399(3): 1191 – 1205.

Cassell, J.; and Tversky, D. 2005. The Language of Online Intercultural Community Formation. *Journal of Computer-Mediated Communication*, 10(2). JCMC1027.

Danescu-Niculescu-Mizil, C.; and Lee, L. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

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

Duff, P. A. 2007. Second language socialization as sociocultural theory: Insights and issues. *Lang. Teach.*, 40(4): 309–319.

Duff, P. A.; and May, S. 2017. *Language socialization*. Springer International Publishing. ISBN 978-3-319-02254-3.

Forte, A.; and Lampe, C. 2013. Defining, understanding, and supporting open collaboration: Lessons from the literature. *American Behavioral Scientist*, 57(5): 535 – 547.

Giles, H. 1973. Accent Mobility: A Model and Some Data. *Anthropological linguistics*, 87–105.

Giles, H.; Coupland, N.; and Coupland, J. 1991. Accommodation theory: Communication, context, and consequence. Contexts of accommodation: Developments in applied sociolinguistics., 1–68. Paris, France: Editions de la Maison des Sciences de l'Homme. ISBN 9780511663673.

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

Jackson, C.; Crowston, K.; Østerlund, C.; and Harandi, M. 2018. Folksonomies to Support Coordination and Coordination of Folksonomies. *Computer Supported Cooperative Work (CSCW)*, 22(2): 1 – 32.

Jackson, C.; Østerlund, C.; Crowston, K.; Harandi, M.; Allen, S.; Bahaadini, S.; Coughlin, S.; Kalogera, V.; Katsaggelos, A.; Larson, S.; Rohani, N.; Smith, J.; Trouille, L.; and Zevin, M. 2019a. Teaching citizen scientists to categorize glitches using machine learning guided training. *Computers in Human Behavior*, 105: 106198.

Jackson, C. B.; Øesterlund, C.; Harandi, M.; Kharwar, D.; and Crowston, K. 2019b. Linguistic Adoption in Online Citizen Science: A Structurational Perspective. Fortieth International Conference on Information Systems, 1 – 17.

Keel, W. C.; Lintott, C. J.; Schawinski, K.; Bennert, V. N.; Thomas, D.; Manning, A.; Chojnowski, S. D.; Arkel, H. v.; and Lynn, S. 2012. THE HISTORY AND ENVIRONMENT OF A FADED QUASAR: HUBBLE SPACE TELESCOPEOBSERVATIONS OF HANNY'S VOORWERP AND IC 2497. *The Astronomical Journal*, 144(2): 66 – 39.

Kraut, R. E.; Resnick, P.; Kiesler, S.; Burke, M.; Chen, Y.; Kittur, N.; Konstan, J.; Ren, Y.; and Riedl, J. 2012. *Building Successful Online Communities*. MIT Press. MIT Press. ISBN 0262297396.

Lam, W. S. E. 2008. *Language Socialization in Online Communities*, 2859–2869. Boston, MA: Springer US. ISBN 978-0-387-30424-3.

Ling, J.; and Jackson, C. B. 2024. Exploring Convergence in Relation using Association Rules Mining: A Case Study in Collaborative Knowledge Production. *arXiv preprint arXiv:2404.15440*.

Lintott, C. J.; Schawinski, K.; Keel, W.; Arkel, H. v.; Bennert, N.; Edmondson, E.; Thomas, D.; Smith, D. J. B.; Herbert, P. D.; Jarvis, M. J.; Virani, S.; Andreescu, D.; Bamford, S. P.; Land, K.; Murray, P.; Nichol, R. C.; Raddick, J.; Slosar, A.; Szalay, A.; and Vandenberg, J. 2009. Galaxy Zoo: 'Hanny's Voorwerp', a quasar light echo? *Monthly Notices of the Royal Astronomical Society*, 399(1): 129 – 140.

Luczak-Roesch, M.; Tinati, R.; Simperl, E.; Kleek, M. V.; Shadbolt, N.; and Simpson, R. 2014. Why Won't Aliens Talk to Us? Content and Community Dynamics in Online Citizen Science. Association for the Advancement of Artificial Intelligence.

Maanen, J. V.; Eastin, J.; and Schein, E. H. 1977. Toward a Theory of Organizational Socialization. Technical report.

Matthews, T.; Mahmud, J.; Chen, J.; Muller, M.; Haber, E.; and Badenes, H. 2015. They Said What? Exploring the Relationship Between Language Use and Member Satisfaction in Communities. 819–825.

Moon, S.; Potdar, S.; and Martin, L. 2014. Identifying student leaders from MOOC discussion forums through language influence. In *Proceedings of the EMNLP 2014 workshop on analysis of large scale social interaction in MOOCs*, 15–20.

Mugar, G.; Østerlund, C.; Hassman, K. D.; Crowston, K.; and Jackson, C. B. 2014. Planet hunters and seafloor explorers: legitimate peripheral participation through practice proxies in online citizen science. *Proceedings of the ACM 2014 conference on Computer Supported Cooperative Work*, 109–119.

Nguyen, D.; and Rosé, C. P. 2011. Language use as a reflection of socialization in online communities. *Proceedings of the Workshop on Language in Social Media LSM*, 76–85.

Niederhoffer, K. G.; and Pennebaker, J. W. 2002. Linguistic Style Matching in Social Interaction. *Journal of Language and Social Psychology*, 21(4): 337–360.

Nov, O.; Arazy, O.; and Anderson, D. 2011. Dusting for Science: Motivation and Participation of Digital Citizen Science Volunteers. In *Proceedings of the 2011 iConference*, 68–74.

Ochs, E.; and Schieffelin, B. 2008. Language Socialization: An Historical Overview. Encyclopedia of Language and Education, 2580 – 2594. Springer, Boston, MA. ISBN 978-0-387-32875-1.

Peters, I.; and Weller, K. 2008. Tag gardening for folksonomy enrichment and maintenance. *Webology*, 5(3): 1–18.

Ponciano, L.; and Brasileiro, F. 2014. Finding Volunteers' Engagement Profiles in Human Computation for Citizen Science Projects. *Human Computation*, 1(2): 245–264.

Rohden, F.; Kullenberg, C.; Hagen, N.; and Kasperowski, D. 2019. Tagging, Pinging and Linking – User Roles in Virtual Citizen Science Forums. *Citizen Science: Theory and Practice*, 4(1): P10008 – 13.

Schieffelin, B.; and Ochs, E. 1986. Language Socialization. *Annual review of anthropology*, 15(1): 163–191.

Simpson, R.; Page, K. R.; and Roure, D. D. 2014. Zooniverse: Observing the World's Largest Citizen Science platform. Proceedings of the 23rd conference on the World Wide Web, 1049 – 1054. ISBN 9781450327459.

Stewart, I.; Chancellor, S.; Choudhury, M. D.; and Eisenstein, J. 2017. #anorexia, #anarexia, #anarexyia: Characterizing Online Community Practices with Orthographic Variation. *arXiv*, cs.CL.

Swol, L. M. V.; and Kane, A. A. 2019. Language and Group Processes: An Integrative, Interdisciplinary Review. *Small Group Research*, 50(1): 3 – 38.

Willett, K. W.; Lintott, C. J.; Bamford, S. P.; Masters, K. L.; Simmons, B. D.; Casteels, K. R. V.; Edmondson, E. M.; Fortson, L.; Kaviraj, S.; Keel, W. C.; et al. 2013. Galaxy Zoo 2: detailed morphological classifications for 304,122 galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 435(4): 2835–2860.

Zevin, M.; Coughlin, S.; Bahaadini, S.; Besler, E.; Rohani, N.; Allen, S.; Cabero, M.; Crowston, K.; Katsaggelos, A. K.; Larson, S. L.; Lee, T. K.; Lintott, C.; Littenberg, T. B.; Lundgren, A.; Øesterlund, C.; Smith, J. R.; Trouille, L.; and Kalogera, V. 2016. Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science. *arXiv*, (6): 064003.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exac-

erbating the socio-economic divide, or implying disrespect to societies or cultures? Yes

(b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes

(c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes

(d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes

(e) Did you describe the limitations of your work? Yes

(f) Did you discuss any potential negative societal impacts of your work? Yes

(g) Did you discuss any potential misuse of your work? NA

(h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes

(i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes

2. Additionally, if your study involves hypotheses testing...

(a) Did you clearly state the assumptions underlying all theoretical results? NA

(b) Have you provided justifications for all theoretical results? NA

(c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA

(d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA

(e) Did you address potential biases or limitations in your theoretical framework? NA

(f) Have you related your theoretical results to the existing literature in social science? NA

(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA

3. Additionally, if you are including theoretical proofs...

(a) Did you state the full set of assumptions of all theoretical results? NA

(b) Did you include complete proofs of all theoretical results? NA

4. Additionally, if you ran machine learning experiments...

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA

(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA

(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...

(a) If your work uses existing assets, did you cite the creators? Yes

(b) Did you mention the license of the assets? NA

(c) Did you include any new assets in the supplemental material or as a URL? NA

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? No

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see **?**)? NA

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see **?**)? NA

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...

(a) Did you include the full text of instructions given to participants and screenshots? NA

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA

(d) Did you discuss how data is stored, shared, and deidentified? NA

Appendix

Example Gravity Spy Discussion Thread

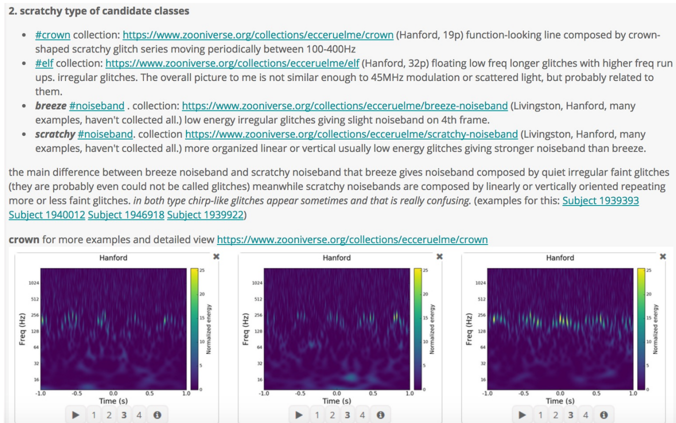

Figure 4: An example discussion page on the Gravity Spy project. The page outlines different types of scratchy glitches such as crown, elf, breeze noiseband, and scratchy noiseband, with links to their respective collections. It also provides a detailed comparison between breeze noiseband and scratchy noiseband glitches. Example subjects and spectrograms illustrating the glitches are included.

A Schematic of Tag Relationships

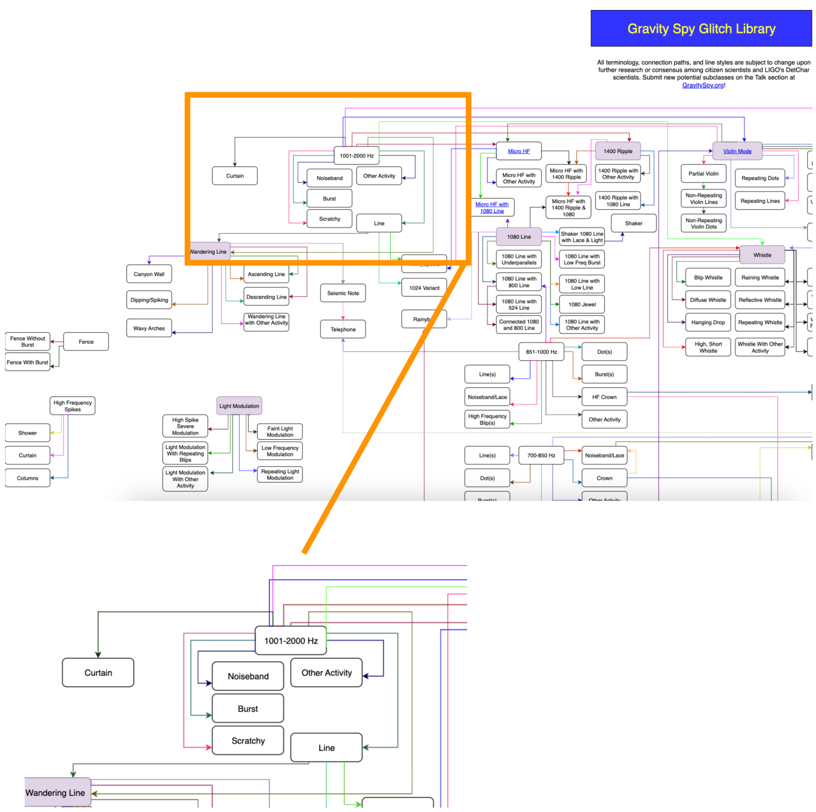

Figure 5: A schematic view of glitches used in Gravity Spy, collected in 2018. A single volunteer created the figure to illustrate the relationships among different glitches identified in the project.