Efficient, but Effective? Volunteer Engagement in Short-Term Virtual Citizen Science Projects

NEAL REEVES and ELENA SIMPERL, Web and Internet Science, University of Southampton

Virtual citizen science (VCS) projects have proven to be a highly effective method to analyse large quantities of data for scientific research purposes. Yet if these projects are to achieve their goals, they must attract and maintain the interest of sufficient numbers of active, dedicated volunteers. Although CSCW and HCI research has typically focussed on designing platforms to support long-term engagement, in recent years a new project format has been trialled – using short-term crowdsourcing activities lasting as little as 48 hours. In this paper, we explore two short-term projects to understand how they influence participant engagement in the task and discussion elements of VCS. We calculate descriptive statistics to characterise project participants. Additionally, using calculation of correlation coefficients and hypothesis testing, we identify factors influencing volunteer task engagement and the effect this has on project outcomes. Our findings contribute to the understanding of volunteer engagement in VCS.

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

Additional Key Words and Phrases: Social Computing; Online Communities; Crowdsourcing; Citizen Science

ACM Reference Format:
Neal Reeves and Elena Simperl. 2019. Efficient, but Effective? Volunteer Engagement in Short-Term Virtual Citizen Science Projects. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 177 (November 2019), 35 pages. https://doi.org/10.1145/3359279

1 INTRODUCTION

Citizen science describes the engagement of volunteers in scientific research, not only through the collection and processing of data for scientific investigation, but in activities such as conservation, community action, raising awareness and education [31, 71, 72]. An increasingly popular form of citizen science is *virtual* citizen science (VCS), a crowdsourced approach conducted almost entirely through virtual portals and covering predominantly data analysis and processing tasks [71]. More so than any other form of citizen science, it is VCS that has seen acceptance from the wider scientific community [74] and which has been applied to a wide range of disciplines and research issues, including astrophysics [26], neuroscience [27] and biology [3] to name but a few. Furthermore, the use of VCS has resulted in a number of major scientific discoveries, such as algorithms for understanding biological molecules [35], new planets [59] and entirely new types of astrophysical phenomena [37].

Yet despite their success, these virtual citizen science (VCS) projects can place large burdens on contributing volunteers. In some VCS projects, quantities of data are far greater than can be analysed by more traditional scientific workflows, which typically rely on individual researchers or

Authors' address: Neal Reeves, neal.reeves@soton.ac.uk; Elena Simperl, e.simperl@soton.ac.uk, Web and Internet Science, University of Southampton, P.O. Box 1212, Southampton, Hampshire, SO1.

2573-0142/2019/11-ART177
https://doi.org/10.1145/3359279

small groups to complete classifications manually [38]. Many of the earliest Zooniverse[1] projects, for example, possessed data sets numbering hundreds of thousands of images for volunteers to complete [68]. This is further compounded by the need to complete quality assurances processes. Web-based crowdsourced citizen science projects tend to make use of aggregation and redundancy to ensure the quality and accuracy of volunteer submissions, which further raises the workload assigned to volunteers [53]. At the same time, there are significant variations in the contribution levels of individual volunteers. While a small minority make the vast majority of contributions, the majority contribute sporadically and for increasingly brief periods [50, 57, 68].

As a result, for projects to successfully achieve their research goals, they must attract and maintain the interest of a sufficiently large community of volunteers – a "critical mass" – throughout the lifespan of the project [53]. The issue of how to sustain this community has been a pervasive question surrounding citizen science research, particularly within the area of CSCW and HCI. A broad range of solutions has been suggested, from the technical such as interface design [64] and gamification [17], to the social such as discussion [66] and even psychological such as motivational factors and affordances [56]. While these techniques have had varying levels of success, the loss of long-term participants is a pervasive issue within VCS projects [57]. At the same time, since this loss occurs at different rates and times within individual projects, it remains difficult to design and account for and causes significant variation in the success and completion rates of projects [8].

One emerging approach within VCS is the introduction of short-term citizen science goals. The Zooniverse platform has begun to develop short-term projects and partnered with the British Broadcasting Company (BBC) and more recently, the Australian Broadcasting Company (ABC) to launch these live on the astrophysics television programme *Stargazing Live* [69]. These projects are carried out over the course of approximately 48 hours, with the aim to complete all necessary classifications for a given project such that results can be unveiled in the third and final live broadcast. This approach has proven highly effective in terms of classifications and scientific discoveries and the initiative has been repeated in consecutive years for a number of astrophysics-related projects [10]. Such an approach marks something of a paradigm-shift for the design of citizen science projects, where it has often been assumed that projects should aim to encourage long-term participation patterns among volunteers if they are to be successful [12]. Nevertheless, short-term challenges have been used to boost participation in established citizen science projects [54] and events such as the BioBlitz are used extensively to encourage data collection for citizen science projects [48].

Beyond increased task engagement, however, is the question of how short-term projects influence concurrent activities in VCS. Projects are increasingly turning to online discussion spaces to engage, train and manage volunteers [24] and volunteers who use such spaces tend to make more contributions than those who do not [39, 68]. Engagement with these spaces is associated with a number of significant benefits for participants serving as an opportunity to learn about the project [20, 39] being in many cases the only space in which feedback on tasks is provided [53]. Discussion spaces offer a degree of autonomy to volunteers which is not seen in task workflows [20, 53] and this in turn can in rare circumstances lead to major scientific discoveries that may have otherwise been overlooked [68]. Yet the benefits of discussion participation are often primarily seen in the long-term [10, 20, 24] and previous analysis has demonstrated that response times for Zooniverse *Talk* discussions often average hours or even days [39], calling into question how suitable *Talk* is for short-term events.

Additionally, there is the question of how these short-term projects perform relative to more conventional VCS projects. The high-level of classifications received by VCS projects in their

[1]The largest platform for VCS projects, see section 4

opening days has been well documented [57]. Do short-term projects truly gather significantly more classifications than those which do not impose such a limit? And how do such projects influence activity in other concurrent projects, particularly in a platform such as the Zooniverse with many simultaneous projects and a potentially finite supply of volunteer effort [8]?

In this paper, we analyse task and discussion contributions from two short-term citizen science projects carried out over the course of 48 hours in February and March 2017. Using correlation analysis, hypothesis testing and measures of effect size, our analysis explores engagement and contribution levels from volunteers, with a particular focus on understanding the impact of this time limit on task and discussion contributions – as well as the resulting relationship between the task and discussion. We then discuss the implications of these findings and situate them within related work from the wider literature and studies of similar, long-term projects. Our paper contributes to the understanding of volunteer engagement patterns within citizen science and aims to suggest novel and effective strategies for implementing crowdsourced methodologies for scientific research.

2 RELATED WORK

Maximising contribution levels [28], recruiting participants [57] and retaining these participants [29, 57] are central research issues surrounding citizen science. Unsurprisingly, therefore, a wealth of research has considered the question of how to design citizen science projects to enhance and encourage participant motivation. Qualitative analyses – particularly interviews with volunteers – have considered the motivation for engaging with discussion and task activity, finding that discussion activity motivates greater levels of participation and supports dabbling behaviour [10, 12]. At the same time, discussion and task elements should not be seen as a dichotomy. In our previous work, we examined the inclusion of online community features in 48 citizen science projects across four categories[2] and found that community features were often embedded within – or otherwise played a significant role in – encouraging task contributions from volunteers [53]. We therefore endeavoured to consider discussion activity alongside task contributions wherever possible.

Further to the embedding of social features within tasks, previous research has considered the relationship between task completion and discussion activity within the Zooniverse and other VCS projects and platforms. Luczak-Roesch et al. [39] analysed 50 million classifications from 250,000 users across 10 Zooniverse projects, exploring discussion activity and discussion-based learning and its correlation with task completion. Similarly, Tinati et al. [66] explored 4,000,000 submissions from 90,000 users within the EyeWire project, exploring the correlation between discussion and task submissions, game length and player churn. Each study found positive correlation between Talk and task activity, although the strength of this correlation varied strongly between projects. Our analysis builds on these metrics, applying them to short-term citizen science, but also goes further in exploring the correlation between Talk and task submissions among individual participants.

Similarly, the work of Jackson et al. [21] analysed the early contribution patterns of citizen science volunteers, distinguishing between specific factors including elements of discussion (Talk) and task contributions. The authors' findings demonstrate a range of clustered citizen science profiles, with a more nuanced association between Talk and task than that suggested by previous work. Rather than a simple linear association between Talk and task activity, different volunteer behaviours – such as viewing discussions and assets within Talk and making comments – are associated with different broad levels of task contribution. We build upon this work in considering how to analyse citizen contributions and in selecting and analysing metrics to understand initial and short-term citizen engagement.

[2]task visibility, goals, feedback and rewards

More specific to task contributions, Sauermann and Franzoni [57] calculated a variety of metrics across seven Zooniverse projects to explore the characteristics of individual contribution and how these contribution patterns vary between participants. Of particular interest to this analysis is the finding of significantly skewed distribution of effort between volunteers and the analysis of the user lifespan. We build on these metrics and partially situate our findings in the context of those of Sauermann and Franzoni [57], although the difference in nature of the studied projects prevents a full like-for-like comparison.

Additionally, further metrics have been suggested by Cox et al. [8], who define a set of success metrics for Zooniverse projects. The authors identified a range of project characteristics that are associated with successful outcomes in projects, across a range of factors such as the completeness of analysis and the distribution of effort among volunteers. Of particular relevance to our research is the consideration of the outcome of project lifespan and how this is associated with other such metrics such as task completion and participant engagement. We draw upon these findings, specifically the metric of distribution of effort and consideration of project lifespans to understand the trade-offs and impact of adding time limits to citizen science methodologies.

The Zooniverse offers volunteers the opportunity to contribute anonymously, without the need to register. This opportunity is important for two key reasons – the first is that individual users often contribute for very brief periods of time and it is important to keep barriers to entry low, to allow projects to make use of as much of a participant's time as possible [50, 68]. Secondly and equally important is the finding that a significant proportion of volunteers benefit from the opportunity to contribute anonymously – even volunteers who register show a tendency to contribute more when projects give the option to register, rather than making registration compulsory [23].

Yet analysis of these anonymous users is difficult. Jackson et al. [22] examined contributions to two Zooniverse projects, distinguishing between those volunteers who logged in to the project and those who contributed anonymously (identified by IP address). The authors found that while a minority, consideration of anonymous users can have import methodological implications for the consideration of citizen science log data. While we were unable to use anonymous traces within our analysis, we drew upon the work of Jackson et al. [22] in aiming to understand how best to consider the impact and importance of anonymous users within the metrics which we calculated.

3 RESEARCH QUESTIONS

3.1 Effectiveness of Short-Term Projects

Sauermann and Franzoni [57] identified that effort in citizen science projects is generally at its highest in the initial days after the project launches, a finding which has also been found in other citizen science projects [54]. Contribution levels during these initial periods may be highly intense, with over five times the level of effort donated by volunteers just three months later [57]. While Curtis [10] notes the success of these short-term projects, within the Planet Four project, the classifications gathered within the initial 48 hour period exceeded those gathered during the next two years combined [3]. This raises the first research question: **How do short-term projects influence the number of classifications gathered from volunteers?**

3.2 Impact on Other Projects

The Zooniverse first launched in 2009 with the Galaxy Zoo project and has since grown to encompass a wide variety of disciplines and tasks [68]. With the launch of the Panoptes service in 2015 it has become increasingly possible for any interested party to design and launch their own Zooniverse project and the number of projects has grown accordingly [61]. Through the use of templates

[3] http://blog.planetfour.org/2015/03/18/2-years-on-from-bbc-stargazing-live/

and a more accessible project design interface, Panoptes removes the complexities associated with designing and implementing citizen science projects, such as the need to program [5]. While volunteers from existing projects have been observed to contribute to new projects in large numbers, causing significant cross-project effects, these effects do not always occur [8, 39]. Moreover, the question is raised of how volunteers spread their effort across projects. Many volunteers contribute only for brief and increasingly sporadic periods and should a project attract insufficient numbers of new volunteers, it will likely exhaust the finite effort that volunteers can contribute to the platform [8, 50]. From here, we derive our second research question: **How do short-term projects impact activity in other concurrent projects?**

3.3 Discussion Activity

Previous studies have suggested a relationship between task completion and discussion activity. Luczak-Roesch et al analysed task and discussion contribution rates across Zooniverse projects and found that volunteers who contributed at least one discussion contribution had higher task completion rates on average than those volunteers who did not, a finding that has also been found in subsequent Zooniverse projects and in the EyeWire platform [21, 39, 66]. Nevertheless, Tinati et al. [67] found only weak correlation between the number of task contributions and the number of Talk comments made by Zooniverse volunteers. Furthermore, analysis suggests that this is a direct result of longer-term contribution patterns in discussion participants, who contribute for longer than the majority of volunteers who do not engage with Talk [21].

This naturally raises the question of whether – given the brief time available to volunteers – volunteers will contribute to Talk or avoid it in favour of completing additional tasks. Our third research question is: **how do short-term projects influence discussion activity in VCS projects?** And similarly, given the suggestion that Talk activity influences task completion, we further consider **what – if anything – is the association between Talk and task in short-term VCS projects?** In particular, we address two key outcomes – an increased user lifespan and an increased number of task submissions.

We note further factors that may fail to influence task and discussion activity in short-term projects. Response times in Talk are significantly higher than in other online community systems, with median response times as high as over 48 hours [39]. Although it is unknown exactly how receiving feedback influences participant behaviours, Talk is the predominant means to garner feedback across Zooniverse projects [53]. Yet it is unlikely that the majority of volunteers will receive a response to their queries or feedback on their findings, despite this being a key driver in the decision to use discussion features [10, 42]. Moreover, in short-term Zooniverse projects, science team members actively participate in Talk – something which is rare in citizen science [53].

We therefore derive eight hypotheses:

- H.1 - Volunteers who contribute to Talk make a greater number of task submissions than volunteers who do not.
- H.2 - Volunteers who contribute to Talk remain with projects longer[4] than volunteers who do not contribute to Talk.
- H.3 - Volunteers who receive at least one reply to a Talk comment make a greater number of task submissions than volunteers who do not.
- H.4 - Volunteers who receive at least one reply to a Talk comment make a greater number of Talk comments than volunteers who do not.
- H.5 - Volunteers who receive at least one reply to a Talk comment remain with projects longer than volunteers who do not.

[4]Here defined as the time between the first and last submission or comment made by a volunteer.

- H.6 - Volunteers who receive at least one reply from a science team member make a greater number of task submissions than volunteers who only receive replies from other users.
- H.7 - Volunteers who receive at least one reply from a science team member make a greater number of talk comments than volunteers who only receive replies from other users.
- H.8 - Volunteers who receive at least one reply from a science team member remain with projects longer than those volunteers who only receive replies from other users.

4 PLANET NINE AND EXOPLANET EXPLORERS

Our analysis draws on two projects from the Zooniverse platform[5], the largest web-based citizen science platform [70].

The first project analysed within this paper is Planet Nine (PN). Planet Nine launched in March 2017 and aimed to find evidence for a hypothesised ninth planet within our solar system. Volunteers were presented with images taken from the SkyMapper digital telescope[6] and asked to classify these according to a single yes or no question, based on the presence of coloured dots within the image (see Figure 1). Despite the similar name, Planet Nine is distinct from the *Backyard Worlds: Planet Nine* project launched earlier on the Zooniverse platform, which features a more complex mapping task and somewhat different targets [30]. As well as launching earlier, Backyard Worlds: Planet Nine continued to run during and after the completion of the Planet Nine project.

Exoplanet Explorers (EE) launched in April 2017 and aimed to find evidence of new exoplanet candidates using light curve graphs generated from data gathered by the *Kepler Space Telescope* [7]. The sheer volume of data generated makes manual classification implausible for classifying the data gathered by the telescope and so a human-in-the-loop crowdsourced approach was used instead. The project builds on the previous success of the similarly focused *Planet Hunters* project, itself the focus of a short-term classification goal, although the project was launched prior to – and continued subsequent to – the launch of this goal [10]. In spite of the brief initial lifespan of the project, volunteers succeeded in finding several potential exoplanet candidates, with the first confirmed discovery from Exoplanet Explorers including 5 planets orbiting one star [7]. As with Planet Nine, volunteers were asked to classify assets according to a single, binary yes/no question, although in Exoplanet Explorers the assets provided to volunteers are light curves, as demonstrated by Figure 2.

In addition to the task interface, each project was accompanied by the Zooniverse discussion platform *Talk*. Talk is a custom-built asynchronous discussion environment, which is based on more traditional discussion forum structures, but designed to address issues of scale and structure which make discussion forums less suited to citizen science contexts [53, 68]. Prompts to engage with Talk are integrated into project task workflows, although the specific format and clarity of these prompts varies between projects [14, 68]. In the case of Planet Nine and Exoplanet Explorers, a link to Talk was shown to each participant prior to the final step of the task submission workflow.

Discussion within the Talk interface takes one of two forms: more traditional forum-based discussions arranged into threads with comment replies and a separate set of "micropost" comments limited to 140 characters and associated with an individual classification asset [45] In the case of Planet Nine and Exoplanet Explorers, three thread categories were available to volunteers: *Science*, *Help* and *Introduce Yourself*. Talk is integrated into task workflows, with prompts to leave comments integrated into the classification process within some Zooniverse projects [68]. Users receive notifications and emails when volunteers reply to their posts within Talk and this can also

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

[6]see https://www.zooniverse.org/projects/skymap/planet-9/about/research

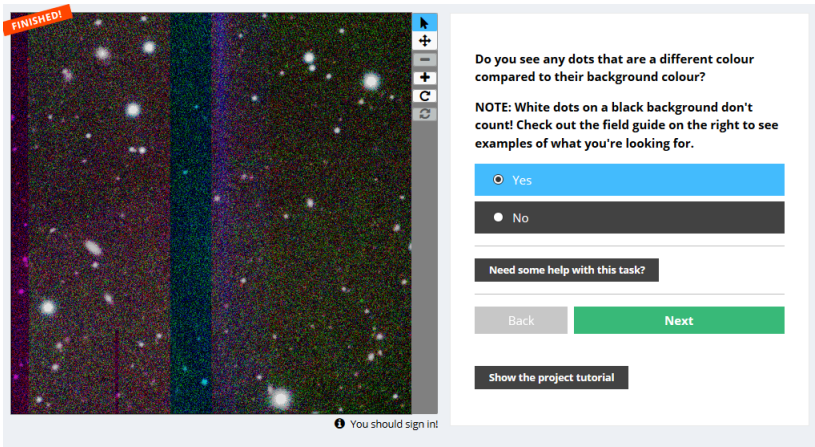

Fig. 1. Interface for Planet Nine showing asset on left and task interface options on right.

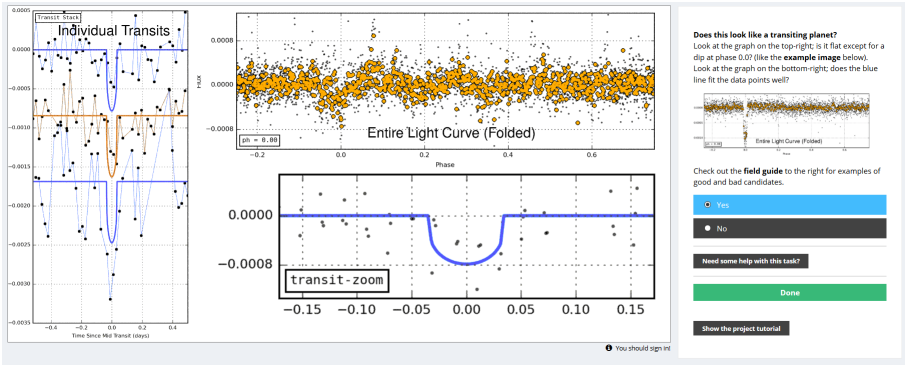

Fig. 2. Interface for Exoplanet Explorers showing asset on left and task interface options on right.

be used to “ping" users through a mention system similar to that of the social network Twitter, although volunteers can modify their notification settings to prevent email notifications [19, 55].

While these discussions have in rare cases led to scientific discoveries [7] participants are equally encouraged to engage in irrelevant or fun discussion activities, such as the sharing of memes [14, 53]. Participation in Talk is entirely optional and it is relatively common for participants to engage with projects without viewing or contributing to Talk [14, 21, 39]. Nevertheless, since task submissions are not publicly visible, Talk is the only way for participants to share contributions with one another and garner feedback from fellow participants and project scientists [53]. Talk further plays an important role in allowing newcomers to learn about tasks from more experienced contributors [20].

5 DATA AND METHODS

We gathered all classifications and all discussion comments made during two sample periods, each corresponding with the duration of one short-term project. The first sample period covers from 7pm GMT on the 28th of March to 3pm GMT on the 3rd of April 2017. This sample contained 5,271,758

[7] see for example [6, 37]

classifications made across 111 projects by 24,689 registered users and a number of anonymous users, as well as 33,704 Talk comments made across 43 projects by 3,408 registered users. The second sample period covers from 4pm GMT on the 3rd of April to 12pm GMT on the 7th of April 2017. During this second sample period, a total of 3,417,808 were made made across 110 projects by 12,838 registered users and a number of anonymous users, with 24,160 Talk comments made across 42 projects by 1,889 registered users.

Each sample period covered the launch and duration of the Planet Nine and Exoplanet Explorers projects respectively, as well as three live broadcast events per sample. It further marks the time during which the three first ever ABC Stargazing Live broadcasts were made, through which Exoplanet Explorers was advertised and results were announced. The difference between the length of the two sample periods is a result of the continued availability of Planet Nine in the days following its informal completion and retirement. While the project was marked as completed, it was still visible within the Zooniverse project interface and available for contribution.

5.1 Inclusion Criteria

For each of the two samples, we drew on all submissions made within the two respective projects – Planet Nine for the first sample and Exoplanet Explorers for the second – and excluded those made in additional projects, except in seeking to characterise participants and when carrying out calculations to understand engagement in concurrent projects. The data used consisted of system log data for the Panoptes task and Talk submission systems and included a number of task submissions made by anonymous or unregistered users. Unfortunately while these submissions could be identified, there was insufficient information within the data to identify which – if any – of these anonymous submissions were made by the same individuals. We therefore removed these anonymous submissions and included only those submissions which were associated with specific usernames or user ID codes.

In line with the findings of Jackson et al. [22], we note that this decision has potential implications for our analysis. Anonymous users were responsible for a minority of submissions in each project – just 7.89% in Exoplanet Explorers and 13.38% in Planet Nine. Moreover, anonymous users cannot contribute to Talk. Given the generally low levels of contribution made by anonymous accounts as described by Jay et al. [23] and Jackson et al. [22], we predict that our calculations surrounding average task completion and hypothesis testing/correlation values are moderately higher than if considering anonymous users. Nevertheless, our findings are not significantly impacted by this variation, as the hypothesis testing figures remain broadly accurate, although the specific effect sizes reported are likely to be lower than if anonymous users were considered.

Talk does not allow for anonymous submissions and so we did not exclude Talk users based on anonymity. Instead, we identified 130 users of Planet Nine who contributed to the Talk element of the project, but from whom no task contributions could be identified. Although we assume this is due to users logging in to make Talk comments while contributing tasks anonymously, we cannot conclusively identify whether these users contributed to tasks or not. We therefore excluded these users from calculations in sections 6.5 and 6.6, to ensure the accuracy of results. Across each sample period, we also removed task and Talk submissions from members of the project science and moderation teams, using these only to detect replies from this group and to distinguish them from other replying users.

5.2 Analysis

We begin our analysis by calculating a number of descriptive statistics, including mean, median, minimum and maximum statistics for each of the two projects. In addition, we present graphical indicators including line graphs and bar charts.

To test the relationship between Talk and task completion in each project, we calculate three distinct measures. The first of these is hypothesis testing. Due to the lack of a normal distribution as observed within each data set, we chose a non-parametric test, the *Mann-Whitney U* test to carry out this process. This particular test was selected due to its lack of assumptions concerning the underlying data, particularly in terms of distribution – while samples used in the Mann-Whitney U test must be from the same population and share a similar distribution, there is no requirement that this be a normal distribution [40], which is essential for this analysis given the power law distribution that contributions follow. An additional advantage of the Mann-Whitney U test is its power and high degree of accuracy (up to 95% of parametric equivalents) in spite of its lack of assumptions [33].

To further understand this relationship, we calculate effect sizes using Cohen's *d*, a commonly reported measure of effect size which we interpret using the updated guidelines of Sawilowsky [58]. Although *d* itself generally assumes normality, we follow the recommendations of Fritz et al. [13] to calculate d from non-parametric equivalents. Finally, we calculate correlation using Kendall's τ, selected due to its reduced bias for very large sample sizes such as those used within this study and reduced vulnerability to false negative rates compared with the similar Spearman's ρ statistic [34, 75].

To understand the impacts short-term projects have on concurrent projects, we additionally analysed the number of classifications and discussion comments made in six 48 hour periods: the 48 hours prior to the launch of Planet Nine and Exoplanet Explorers, the 48 hours during which the classification challenges took place and the 48 hours after the challenge was complete. This data was sampled from the 10 projects which received the most activity (classification and comment levels) during the sampled period, representing approximately 25% of the official Zooniverse projects which were available during this time[8]. Although this analysis considers the project challenge period only, it should be noted that projects remained available after this point and participants continued to contribute for several hours, albeit without the pressure of the associated time limit.

We then used the nonparametric Kruskal-Wallis H test to identify whether a statistical significant difference was present in the number of classifications and discussion comments made during these periods. Where a statistically significant difference was identified, Dunn's non-parametric comparison test was conducted to identify specific pairwise results [11]. A post-hoc Bonferroni correction for multiple comparisons was conducted for each Dunn's non-parametric comparison test.

6 RESULTS

Between March the 28th and the 3rd of April, a total of 5,271,758 classifications were made within the Zooniverse of which 4,369,620 (82.89%) of classifications were made within the Planet Nine project. In addition, 33,704 Talk comments were made across the Zooniverse during this time, of which 20,346 comments were made within Planet Nine (60.37%). A total of 111 projects received at least one classification during this time, of which 43 were fully launched, official Zooniverse projects, with the remaining projects Panoptes project builder and beta projects. Similarly, 42 projects received at least one Talk comment, of which 40 were fully launched projects. During the second sample period, a total of 3,417,808 classifications were made of which 2,762,043 classifications (80.81%) were made in Exoplanet Explorers alone, while 24,113 Talk comments were made of which 17,844 (74.00%) were made within Exoplanet Explorers. In total, 110 projects received at least one contribution during this time, while 43 projects received at least one Talk comment. Descriptive

[8]Beyond these 10 projects, hourly activity levels were highly sporadic and unpredictable and identifying an average level of classifications was not possible

Table 1. Descriptive statistics for Exoplanet Explorers and Planet Nine, calculated based on registered users only.

Statistic	Planet Nine	Exoplanet Explorers
Mean Task submissions	187.01	311.67
Median Task submissions	58	99
Mean Talk comments	8.67	14.20
Median Talk comments	5	5
Mean User Lifespan (hh:mm:ss)	12:35:01	13:07:40
Median User Lifespan (hh:mm:ss)	1:08:28	0:49:59
Min User Lifespan (hh:mm:ss)	00:00:00	00:00:00
Max User Lifespan (hh:mm:ss)	139:38:54	92:10:06
Total Task submissions	3,784,891	2,544,448
Total Talk comments	23,355	17,844
Number of Threads	181	194
Number of Thread Posts	822	597
Number of *Notes* Comments	23,050	17,247
Total Volunteers	20,346	8,164
Talk Participants	2,693	1,251

statistics for registered users in Planet Nine and Exoplanet Explorers can be seen in table 1. A total of 20,346 users contributed to Planet Nine, while a smaller 8,164 contributed to Exoplanet Explorers. While the full number of anonymous users is unknown, registered users completed the majority of tasks in both Planet Nine (86.62%) and Exoplanet Explorers (92.12%).

Talk participation in both projects generally took the form of micropost-style comments on the "Notes" board, with 98.69% of Planet Nine Talk contributions and 96.65% of Exoplanet Explorers Talk contributions taking this form. Nevertheless, it should be noted that in both Planet Nine and Exoplanet Explorers, these comments appear twice for each asset – once on the asset page itself and a second time in the "Notes" category, which is part of the more traditional thread and board style. Help was the most popular forum category in each project (EE: 83 threads; P9: 112 threads) followed by Science (EE: 83 threads; P9: 41 threads) and finally Introduce yourself (EE: 32 threads; P9: 28 threads).

6.1 Characterising Participants

The Zooniverse platform does not request demographic data on individual participants and both Planet Nine and Exoplanet Explorers are no exception to this rule. Based on log data, however, we are able to characterise participants according to location and registration date.

6.1.1 Location. A substantial majority of registered Planet Nine volunteers were based in the UK (13,922, 68.65%), followed by the US (2,334, 11.51%) and Australia (2.85%). When considering only Talk participants, the proportions of volunteers and countries they contributed from was broadly similar, with a majority from the UK (1,667, 65.2%), followed by the US (330, 12.92%) and a third significant group from Australia (108, 4.23%). Conversely, in Exoplanet Explorers – broadcast in Australia – the largest proportion of users (3,333, 41.02%) contributed from Australia, followed by the UK (1,549, 19.06%) and then the US (17.23%). Similar results are seen in Talk, where the largest proportion of users contributed from Australia (471, 37.89%), followed by the US (220, 17.70%) and the UK (197, 15.85%). Within each project, the largest proportion of volunteers stemmed from the country in which the broadcast took place, suggesting broadcasts impact the likelihood that volunteers will contribute.

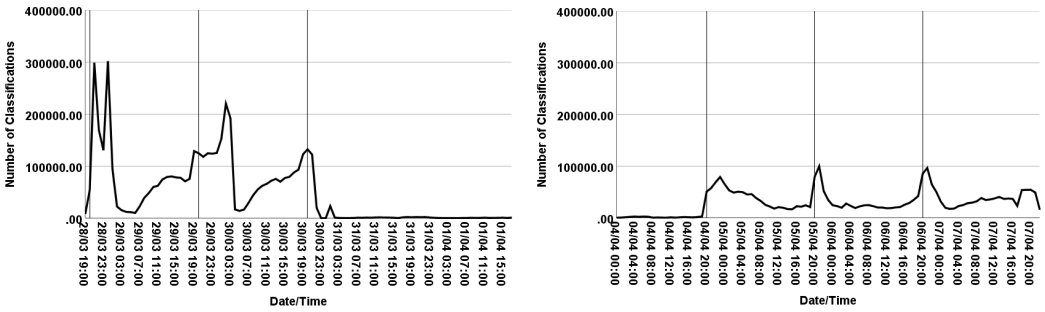

Fig. 3. Line graph showing hourly contribution counts within the Planet Nine (left) and Exoplanet Explorers (right) projects. Vertical lines indicate start time of hour-long broadcast events.

6.1.2 Registration Date. In the case of Planet Nine, 11,669 of 20,346 users (57.35%) registered with the Zooniverse after to the launch of the Planet Nine project. The earliest registration was carried out on the 9th of July 2007, while the last registration captured during the sampled period was made on the 3rd of April, 2017. Conversely, in the case of Exoplanet Explorers, the earliest registration was carried out on the 11th of July 2007, while the last registration was captured on the 7th of April, 2017. Of the 8164 volunteers who contributed to Exoplanet Explorers, 3,168 (38.80%) registered with the Zooniverse after the launch of the project.

When examining median registration dates for each project, it becomes clear that the majority of volunteers in each project had only recently signed up to the Zooniverse. The median registration date for Planet Nine was 7:56 pm on the 28th of March 2017, less than an hour after the launch of Planet Nine at approximately 7:00 pm on the 28th of March. Conversely, the median registration date for Exoplanet Explorers – although slightly more recent at 8:11 pm on the 28th of March 2017 – was almost 7 days prior to the soft launch[9] of Exoplanet Explorers and 8 days prior to the beginning of the classification campaign. Talk participants had moderately older account ages when compared with non-Talk participants, with a median registration date of 5:18 pm on the 28th of March 2017 for Planet Nine and 4:32 pm on the 24th of March 2017 for Exoplanet Explorers.

6.2 Contributions over Time

In each project, peak classification and discussion levels coincide with Stargazing Live broadcast events as shown by Figures 3 and 4. In the case of Planet Nine, the impact of these events is reduced with each successive broadcast, with the initial broadcast event generating 300,000 hourly classifications, the second leading to approximately 220,000 hourly classifications and the third and final broadcast generating just 130,000 classifications. In Planet Nine, a similar phenomenon can be seen in the number of users contributing to each project for the first time after each broadcast. While the initial broadcasts drive large numbers of users to contribute to make their first contribution to the project, this effect is significantly reduced for the second and third broadcast, with the number of new users observed following the second broadcast being less than half that of the first.

Conversely, in Exoplanet Explorers there are a number of users who make their first contribution to the project outside of the broadcast cycle, with a significant proportion who contribute for the first time in the hours following the final broadcast event – i.e., the end of the short-term classification event – as demonstrated by Figures 5 and 6. The volume of new task contributors are approximately equal in the hours following the second broadcast and the hours following the

[9]Exoplanet Explorers formally went live at apprximately 8:00 pm GMT on the 3rd of April 2017, but was not publicised until the beginning of the broadcast event the following day

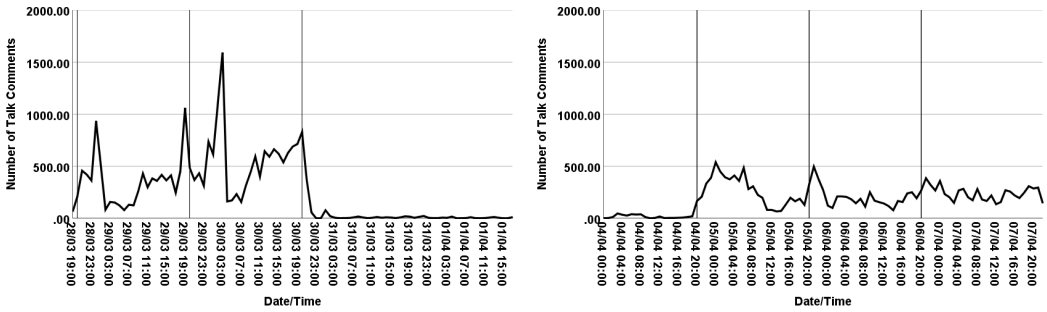

Fig. 4. Line graph showing hourly Talk comment counts within the Planet Nine (left) and Exoplanet Explorers (right) projects. Vertical lines indicate start time of hour-long broadcast events.

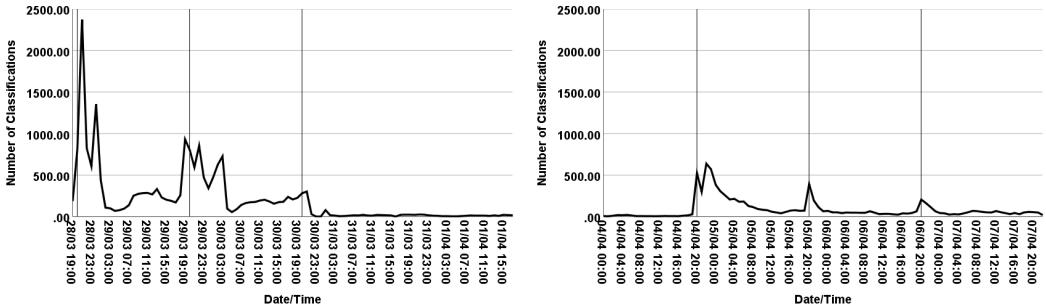

Fig. 5. Line graph showing first classifications made by users within the Planet Nine (left) and Exoplanet Explorers (right) projects.

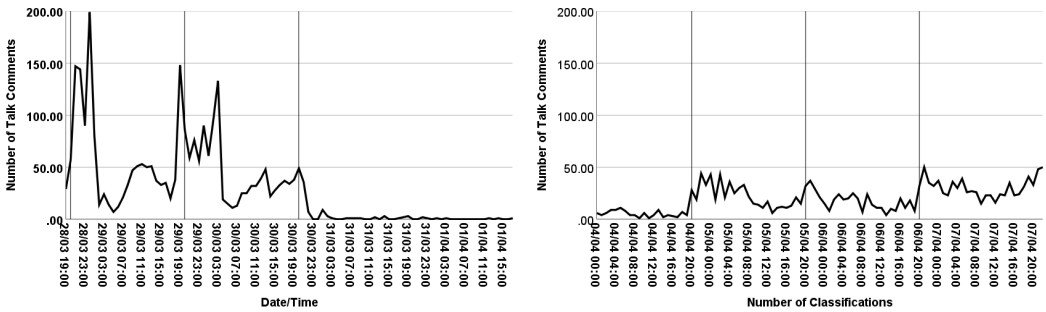

Fig. 6. Line graph showing first Talk comments made by users within the Planet Nine (left) and Exoplanet Explorers (right) projects.

final broadcast and by extension, end of the 48 hour challenge period. Moreover, the number of users making their first comment is at its greatest in the hours following the final broadcast and at the end of the sample period, although no accompanying rise is visible in the number of first-time classifying users. This is not the case in Planet Nine, where there are almost no new task and Talk contributors after the completion of the third broadcast.

Conversely, the impact of broadcasts on discussion is more nuanced across the two projects. In each case, Talk activity is more sporadic and prone to sharp variation, although changes in Talk

Table 2. Launch date, total classifications made during the initial 48 hours and median, lower quartile and upper quartile statistics for hourly classification count during the initial 48 hours for Zooniverse projects launched between January and June 2017.

Project	Launch	Total	Median	Q1	Q3
Astronomy Rewind	22/3/17	45,827.00	471.00	170.25	942.50
Backyard Worlds: Planet Nine	15/2/17	475,724.00	9,226.50	4,839.25	14,725.75
Bash the Bug	7/4/17	14,361.00	262.00	171.50	390.75
Colorado Corridors	2/6/17	13,525.00	272.00	180.00	350.50
Count Flowers for Bees	10/5/17	16,643.00	234.50	82.00	456.00
Elephant Expedition	10/5/17	66,764.00	1,363.00	568.50	2,009.50
Etch a Cell	6/4/17	2,126.00	39.00	19.50	53.75
Galaxy Nurseries	31/5/17	89,769.00	1,428.00	879.75	1,810.50
Michigan ZoomIN	3/5/17	81,633.00	1,466.00	784.00	1,971.00
Muon Hunters	28/2/17	407,488.00	8,489.33	5,188.50	11,781.75
Planet Four: Ridges	17/1/17	203,484.00	3030.50	1,734.00	4,303.00
Plastic Tide	26/4/17	139,336.00	1,488.50	715.25	4,195.00
Steller Watch	15/3/17	245,868.00	3,653.50	547.25	5,727.25
Western Montana Wildlife	12/4/17	77,253.00	1,164.00	709.50	1,604.50
Exoplanet Explorers	4/6/17	1,663,459.00	24,625.00	20,846.75	48,243.25
Planet Nine	28/3/17	4,074,437.00	75,095.50	40,434.50	122,520.50

activity broadly follow changes in task completion. Surprisingly, however, although participants are driven to make their first Talk comments in the hours following live broadcasts, the majority of Exoplanet Explorers volunteers make their first comments in the hours following the completion of the broadcast and the completion of the project, with the number of new Talk participants rising as the number of Talk comments falls. No such influx of Talk participants is seen in Planet Nine, although it should be highlighted that Exoplanet Explorers encouraged participants to mark potential planetary candidates through Talk in a way that Planet Nine did not.

Moreover, as each figure shows, upon 'completion' of the broadcast cycle, participation in each project drops off sharply, within 4 hours for Planet Nine and 8 hours for Exoplanet Explorers. One potential factor in this reduction in activity is the presence of indicators within the projects that the project had 'finished'. Each project featured a message to volunteers thanking them for their contributions and noting that the projects were temporarily completed and did not require further contributions. Additionally, a red banner was added to the corner of each project indicating completion[10]. Nevertheless, Talk participation diminished at a similar rate and similar time to discussion classification, despite no such applicable completion of discussion activity. In fact, both projects demonstrate strong positive correlation between hourly Talk and task contribution rates (τ = 0.70 for both projects).

6.3 Comparison with Other Projects

We analysed the volume of classifications and Talk comments produced in the first 48 hours of 14 projects launched within 3 months of the launch of Planet Nine. Descriptive statistics regarding classifications within each project can be seen in table 2, while Talk comments can be seen in table 3. Figures from Exoplanet Explorers and Planet Nine represent contributions made during the 48 hour short-term project period.

[10]This banner can be seen in the corner of Figure 1

Table 3. Launch date, total talk comments made during the initial 48 hours and median, lower quartile and upper quartile statistics for hourly talk comment count during the initial 48 hours for Zooniverse projects launched between January and June 2017.

Project	Launch	Total	Median	Q1	Q3
Astronomy Rewind	22/3/17	631.00	7.00	2.00	59.75
Backyard Worlds: Planet Nine	15/2/17	2,432.00	47.00	37.25	59.75
Bash the Bug	7/4/17	80.00	0.50	0.00	2.00
Colorado Corridors	2/6/17	68.00	0.00	0.00	2.00
Count Flowers for Bees	10/5/17	56.00	0.00	0.00	1.00
Elephant Expedition	10/5/17	111.00	1.00	0.00	4.00
Etch a Cell	6/4/17	20.00	0.00	0.00	1.00
Galaxy Nurseries	31/5/17	973.00	16.00	4.25	26.00
Michigan ZoomIN	3/5/17	716.00	10.50	5.00	21.75
Muon Hunters	28/2/17	855.00	15.00	7.25	27.75
Planet Four: Ridges	17/1/17	843.00	13.00	7.00	23.50
Plastic Tide	26/4/17	319.00	5.00	1.00	11.75
Steller Watch	15/3/17	1,202.00	13.50	3.00	35.25
Western Montana Wildlife	12/4/17	454.00	6.00	3.25	15.00
Exoplanet Explorers	4/6/17	11,071.00	197.00	142.50	320.50
Planet Nine	28/3/17	20,967.00	406.00	215.50	592.50

Most notably, contributions to Planet Nine and Exoplanet Explorers are an order of magnitude greater than in other projects launched during the same time period. Exoplanet Explorers and Planet Nine each garnered millions of classifications – 1,663,459 for Exoplanet Explorers and 4,074,437 for Planet Nine – while the most popular long-term project launched during this time (Backyard Worlds: Planet Nine) gathered just 475,724 classifications in its first 48 hours. This is equally true of the number of Talk comments made within these projects, with 11,071 comments made in Exoplanet Explorers and 20,967.00 comments made in Planet Nine, compared to just 2,432 comments made in the next highest project – Backyard Worlds: Planet Nine. Ultimately these figures suggest that participation levels in short-term projects exceed the temporary intense contribution levels of VCS project launches.

Nevertheless, a number of factors are associated with motivation and participation in VCS projects. Notable factors include – but are certainly not limited to – task difficulty and associated completion time, the characteristics and contents of the images provided for classification, volunteers' individual motivations and the provision of feedback [43, 53, 68]. While factors such as task difficulty are inherently subjective, Planet Nine and Exoplanet Explorers feature a much less complex task than other projects launched around the same time, with just a binary classification interface and a simple question that can be completed in just a few seconds. Conversely, Etch a Cell asks volunteers to draw around the nucleus of a cell using their mouse pointer and this is a much more time-consuming and complex task. Similarly, Colorado Corridors, Michigan ZoomIN and Western Montana Wildlife ask volunteers to classify animals present in camera trap images and therefore include significantly more than two options from which volunteers must choose. Colorado Corridors, for example, includes 26 different animal species as classification options.

Moreover – and perhaps most crucially – unlike Planet Nine and Exoplanet Explorers, none of the other projects launched during this period were featured in live broadcast campaigns. Although contribution levels in short-term projects *can*, under the right cirumstances exceed those of other more long-term projects, we cannot conclusively identify to what extent this was influenced by

live broadcasts and effective publicity campaigns, nor whether such observations would be seen across other projects, task types and scientific domains.

6.4 Impact on Other Projects

Due to the number of observations and projects involved, full Kruskal Wallis test statistics can be found in Appendix A in Table 6. Graphs showing individual and median hourly classification and talk contributions, as well as the Dunn's non-parametric comparison test outcomes for statistically significant findings can be seen in graphs 7 - 22 in Appendix B.

6.4.1 Planet Nine – Classification Levels. We find some evidence to suggest that the launch of short-term projects may impact engagement in concurrent projects. When comparing the 48 hours prior to, during and after the launch of Planet Nine, the Dunn's test found statistically significant differences in classification levels within 6 projects. In the Backyard Worlds: Planet Nine project, which shares a number of similarities with the Planet Nine short-term project, our findings suggest that Planet Nine may have negatively impacted classification levels within the project. When comparing the 48 hours prior to the launch of Planet Nine with the 48 hours during which the short-term classification challenge occurred, we find a statistically significant difference ($p<0.001$) in classification levels and lower median classification levels during the course of Planet Nine (1047.50) than before Planet Nine launched (2670.50). Similarly, when comparing the duration of the project with the 48 hours subsequent to the end of the challenge, the test outcomes suggest a significant difference ($p<0.001$) and higher median classifications after the completion of Planet Nine (2156.50).

Similarly, in the Comet Hunters project, median classification levels were higher in the 48 hours prior to the launch of Planet Nine than in the 48 hours during the project (88 vs 77) and highest in the 48 hours subsequent to the completion of the project (154). Dunn's test outcomes show a statistically significant difference ($p<0.001$) in classification levels between the 48 hour period prior to and 48 hour period subsequent to the project, as well as the 48 hour period during and 48 hour period subsequent to Planet Nine ($p=0.012$).

In other projects, our findings suggest a positive impact on project engagement levels. Test outcomes show a statistically significant difference ($p=0.001$) between classification levels in the 48 hours prior to and the 48 hours during the Gravity Spy project, with higher median classification levels during Planet Nine (465.00 vs 243.50). Moreover, classification levels were significantly different ($p<0.001$) when comparing the 48 hour periods prior to and after Planet Nine, with higher median classification levels after Planet Nine (med=480.00 vs 243.50). In the Steller Watch project too, median classification levels were highest during the course of Planet Nine, with a statistically significant difference ($p=0.007$) in classification numbers when comparing the course of the project with the 48 hour period subsequent to its completion (med=1030.50 vs 706.50). Conversely in Muon Hunters, median classification levels were at their lowest during the course of Planet Nine, with a statistically significant difference ($p=0.040$) in classifications when comparing the duration of Planet Nine with the 48 hours subsequent to its completion (med=565.50 vs 793.50).

A third group of projects show a consistent change in classification levels over time. In Astronomy Rewind, median classification levels were highest prior to the launch of Planet Nine at 210, falling to 141 during the launch of the project and 119 in the 48 hours after the project was completed, with a statistically significant difference ($p=0.030$) between the 48 hours prior to and 48 hours during the project, as well as the 48 hours prior to and 48 hours subsequent to Planet Nine ($p=0.002$).

6.4.2 Exoplanet Explorers – Classification levels. Our findings suggest that the launch of Exoplanet Explorers had less of an impact on activity within the Zooniverse than the launch of Planet Nine, perhaps linked to the lower number of participants who engaged with EE. Only two of the sampled

projects recorded a statistically significant variation in classification levels around the launch of EE – the astrophysics themed Backyard Worlds: Planet Nine project and the ecology themed Cleveland Metroparks project. A statistically significant variation ($p<0.001$) was observed within BW: P9 between the 48 hours prior to (med=2150.00) and subsequent to (med=1127.00) – as well as a significant difference ($p=0.002$) between the 48 hours during (med=1510) and 48 hours subsequent to Exoplanet Explorers – but not between the 48 hours prior to and 48 hours during the running of EE.

Conversely, in the Cleveland Metroparks project, a significant difference in classification levels occurred between the 48 hours prior to and 48 hours during Exoplanet Explorers project, as well as a significant difference ($p=0.006$) in classification levels between the 48 hours prior to and 48 hours subsequent to Exoplanet Explorers, but not when comparing the period during the project and the period after the project took place.

6.4.3 Discussion Comments. The Dunn's test outcomes for discussion levels – perhaps surprisingly – demonstrate that there was not necessarily an overlap between those projects in which classification levels varied and those projects in which the number of Talk comments varies. The Camera CATalogue, Chicago Wildlife Watch and Milky Way projects all featured significant variations in discussion levels but not classification levels, while the Astronomy Rewind, Cleveland Metroparks, Comet Hunters and Muon Hunters projects all conversely featured variations in classification level but not discussion level.

The number of discussion comments made within Backyard Worlds: Planet Nine was significantly different only when comparing the period during and period after the project lifespan for Planet Nine, with a lower median number of comments during the project (13.50) compared with after (20.00). Comment numbers were significantly different for Exoplanet Explorers both when comparing the period prior to (med=22.00) and period during (median 17.00) as well as the period during and period subsequent to (med=10.50) the lifespan of the project. This suggests a sustained fall in participation which may be unrelated to Exoplanet Explorers. The opposite phenomenon was observed in Chicago Wildlife Watch, with Talk rising during PN (med=1.00) and after PN (med=2.00) when compared with before the project (med=1.00). In other projects we find little evidence of a significant difference related specifically to the short-term project launch period.

6.5 Distribution of Effort

As with many online collaborative communities, citizen science projects exhibit high levels of inequality between the amount of effort contributed by individual volunteers. Neither Exoplanet Explorers nor Planet Nine were exceptions to this rule. Both projects relied heavily on a minority of highly active volunteers who completed thousands of classifications each, far in excess of the majority of users. Within Planet Nine, the most active participant contributed 14,601 classifications within 48 hours, while in Exoplanet Explorers, the most active participant contributed 11,921 classifications. This is significantly higher than the median average in each project of 58 classifications within Planet Nine and 99 classifications within Exoplanet Explorers. The projects differ, however, in terms of which types of user are most active. In Exoplanet Explorers, the most active classification contributors were also active in Talk, in line with findings from other projects within the Zooniverse [39] and external projects such as the Game With A Purpose EyeWire [66]. On the other hand, in Planet Nine, the most active contributor contributed only to the task element of the project, making over 4,000 more classifications than the most active talk participant.

A similar significant distribution is seen when considering Talk comments. In Exoplanet Explorers, for example, the most active volunteer made 689 Talk comments while in Planet Nine, the most active volunteer made 864 comments, compared to a median of just 2 comments in both projects.

Despite this, there is only a weak correlation between the number of Talk comments made and tasks contributed by users across each project. In Planet Nine, the correlation coefficient τ is equal to 0.30 when considering only those volunteers who contributed to both Talk and task, but just 0.21 when including those volunteers who did not make Talk comments. Similarly in Exoplanet Explorers, the coefficient τ is equal to 0.29 when considering only those volunteers who contributed to both Talk and task , falling to just 0.22 when including those volunteers who did not make Talk comments.

6.6 Relationship Between Talk and Task

To test the hypotheses set out previously, we used the Mann Whitney-U hypothesis test to identify if statistically significance differences were present between distinct samples from each of the two projects. A full summary of the tests carried out, test statistics and outcomes, as well as effect sizes can be seen in table 4 for Planet Nine and 5 for Exoplanet Explorers.

In spite of the high levels of task only participation in Planet Nine, hypothesis test *H1* reveals that volunteers who engage with Talk complete more tasks on average than volunteers who do not ($p<0.001$, $d=0.52$), with a medium effect size similar to Exoplanet Explorers. This figure aligns with studies of projects both of Zooniverse projects and external projects that Talk participants contribute more tasks than non-Talk participants [21, 39, 66]. Nonetheless, this finding underpins the importance of Talk even for short-term participation and shows that Talk participants are highly active, rather than merely sporadic but long-term project participants.

In both projects, participants who received at least one reply to their Talk comments made significantly more task contributions ($p<0.001$) overall than those Talk participants who did not receive a reply, with a large effect size in the case of Planet Nine ($d=0.55$) and a medium effect size in the case of Exoplanet Explorers ($d=0.55$). In both projects, this effect extended to the number of Talk comments left by participants, which was greater among those who received at least one reply ($p<0.001$ for both projects). Nevertheless, this effect was much stronger for Exoplanet Explorers ($d=1.15$), than for Planet Nine ($d=0.46$). It is arguably possible that this is simply a matter of probabilities – if each comment is equally likely to receive a reply, then the more comments an individual makes, the more likely it is that he/she will receive a reply. However, the effect size for Exoplanet Explorers – being so large – suggests otherwise.

It should be stressed that response times varied between different thread and discussion categories. In both project, the Notes category had the largest response time, with a mean time of 7 hours 18 minutes for Planet Nine and 7 hours and 40 minutes for Exoplanet Explorers. Forum-based threads had a much shorter response time with Help having a mean response time of 12 minutes for P9 and 37 minutes for EE, Science having a response time of 13 minutes for P9 and 24 minutes for EE and Introduce Yourself having the longest response time in each case of 19 minutes for Planet Nine and 1 hour 19 minutes for Exoplanet Explorers.

This increased level of participation among Talk participants is likely linked to the finding that in both projects, volunteers who engage with Talk have a longer average lifespan than volunteers who only contribute to task elements ($p<0.001$ for both projects). In both projects, the effect size is medium, although in the case of Exoplanet Explorers, the effect size is greater ($d=0.64$) than in Planet Nine ($d=0.50$). In both projects, participants exhibit greater user lifespans if they received at least one reply to a Talk comment ($p<0.001$; PN - $d=0.52$; EE - $d=0.40$). There is, however, a difference in effect size for each project.

One significant difference in the sampled short-term projects compared with other VCS activities is the tendency for project scientists to actively engage with the community through Talk – something which is relatively rare in VCS generally [53]. Our findings suggest that replies from project science team members have a statistically significant effect on user lifespans in both projects,

Table 4. Summary of hypothesis, *H* Mann-Whitney U test outcomes for Planet Nine, where H - hypothesis, Med 1 - median 1, Med 2 - median 2, Z - Z-score, p - p-value and *d* - effect size Cohen's *d*.

H	Condition	Med 1	Med 2	Z	p	*d*	Conclusion
1	Number of Task Submissions (Talk and Task vs Task Only)	188	48	-35.60	<0.001	0.52	Talk participants contributed more tasks on average than task only participants (medium effect size)
2	User Lifespan (Talk and Task vs Task Only - Planet Nine)	19h30	0h38	-34.78	<0.001	0.50	Talk participants remained with projects longer than non-Talk participants (medium effect size)
3	Number of Task Submissions (Received Reply vs No Reply)	219.50	97.50	-18.54	<0.001	0.77	Participants who received a reply to their Talk comments contributed more tasks on average than those who did not (medium effect size)
4	Number of Talk Comments (Received Reply vs No Reply)	2.00	1.00	-11.47	<0.001	0.46	Talk participants who received at least one reply to a Talk comment made more comments on average than those who did not (small effect size)
5	User Lifespan (Received Reply vs No Reply)	22h32	4h05	-10.05	<0.001	0.40	Talk participants who received at least one reply had greater average lifespans than participants who did not receive a reply (small effect size)
6	Number of Task Submissions (Reply from Scientist vs Other Participants)	355	201	-6.14	<0.001	0.28	Talk participants who received at least one reply from a science team member made more task submissions on average than those who only received responses from other participants. (small effect size)
7	Number of Talk Comments (Reply from Scientist vs Other Participants)	1	2	-12.49	<0.001	0.58	Talk participants who received at least one reply from a science team member made more Talk comments on average than those who only received responses from other participants (medium effect size)
8	User Lifespan (Reply from Scientist vs Other Participants)	28h21	19h31	-8.26	<0.001	0.37	Participants receiving reply from scientists had greater average lifespan than participants receiving reply from other participants (small effect size)

but the effect size is small for each project, calling into question how significant this effect is (PN - d = 0.30; EE - d=0.37). This mirrors an overall trend within these results that Exoplanet Explorers volunteers' task contribution patterns appear to be less influenced by engagement with project scientists unlike Planet Nine volunteers. As shown in table 5, although there was a statistically significant difference (p=0.008) in the number of task submissions made by volunteers receiving replies from the science team, the effect size was very small (d=0.18). This is perhaps surprising given that Exoplanet Explorers was successful in finding evidence of exoplanets, while Planet Nine found no evidence for the hypothesised ninth planet. Perhaps, then, it is not indicators of success which influence participant interactions with scientists, but specifically what factors *do* influence these interactions is unclear and requires further study.

Table 5. Summary of hypothesis, *H* Mann-Whitney U test outcomes for Exoplanet Explorers, where H - hypothesis, Med 1 - median 1, Med 2 - median 2, Z - Z-score, p - p-value and *d* - effect size Cohen's *d*.

H	Condition	Med 1	Med 2	Z	p	*d*	Conclusion
1	Task Submissions (Talk and Task vs Task Only - Exoplanet Explorers)	298.50	78.00	-23.61	<0.001	0.54	Talk participants contributed more tasks on average than non-Talk participants (medium effect size)
2	User Lifespan (Talk and Task vs Task Only - Exoplanet Explorers)	23h16	0h28	-27.71	<0.001	0.64	Talk participants remained with projects longer than non-Talk participants (medium effect size)
3	Task Submissions (Received Reply vs No Reply)	370	158	-9.29	<0.001	0.55	Participants who received a reply to their Talk comments contributed more tasks on average than those who did not (medium effect size)
4	Talk Comments (Received Reply vs No Reply)	4	1	-17.64	<0.001	1.15	Participants who received a reply to their Talk comments contributed more talk comments on average than those who did not (large effect size)
5	User Lifespan (Received Reply vs No Reply)	26h32	3h24	-8.83	<0.001	0.52	Talk participants who received at least one reply remained with projects longer than participants who did not receive a reply (medium effect size)
6	Task Submissions (Reply from Scientist vs Other Participants)	362.00	484.50	-2.67	0.008	0.18	There is a statistically significant difference between the number of tasks completed by those who received at least one reply to a Talk comment from the science team and those who only received responses from other participants (very small effect size)
5	Talk Comments (Reply from Scientist vs Other Participants)	4.00	10.50	-6.13	<0.001	0.41	Talk participants who received at least one reply from a science team member made more Talk comments on average than those who only received responses from other participants (small effect size)
8	User Lifespan (Reply from Scientist vs Other Participants)	25h14	46h16	-4.48	<0.001	0.30	Talk participants who received at least one reply from a science team member remained with projects longer than participants who only received responses from other participants (small effect size)

It should be noted that in analysing replies, we did not distinguish between the more forum-based thread-type structure and the subject-specific micropost comments. We note a large number of comments which responded to other comments on the same asset – for example "red dot?" "not sure". This behaviour was displayed by users and scientists alike, but it is unclear whether this is an evolution in the use of Talk or merely a characteristic of Planet Nine, Exoplanet Explorers and potentially short-term projects generally.

7 DISCUSSION

In spite of the similar time limit and use of broadcasting in both projects, our results show that Planet Nine generated significantly more activity than Exoplanet Explorers. We note two key factors that likely influence interest in each project. The first factor is familiarity – while the Zooniverse partnered with Stargazing Live annually between 2013 and 2017 [10], this partnership has typically only resulted in the launch of one project. In contrast, 2017 marked the first time that a second project was launched and moreover, the first time in which ABC Stargazing Live was broadcast. Additionally, the second factor is similarity with existing projects. Zooniverse projects exhibit strong cross-project effects in which participants from existing projects try out new projects which are similar to the existing projects in which they engage [39]. In February 2017, the Zooniverse launched Backyard Worlds: Planet 9, a project which shares many similarities with – yet is distinct from – Planet Nine and which proved highly popular [30]. The Planet Nine project, then, had an existing community of active users within the Zooniverse platform and our results show that 36.27% of PN users had also contributed to BW:P9.

Nevertheless, in spite of the similarity in name and purpose, no formal cross-project marketing or collaboration took place between Backyard Worlds: Planet 9 and the Stargazing Live Planet 9 projects. No Talk comments were made to announce the launch of a new project and no BW:P9 blog posts were made about the Stargazing Live activity. BW:P9 contributors would have been made aware of P9 through an email upon the launch of the project which may have spurred this strong cross-project effect, but this email was sent to all registered Zooniverse users and was not project specific.

7.1 Impact on Volunteer Engagement

A common finding among existing Zooniverse projects is that the majority of contributions occur in the earliest days of the project and these contributions diminish over time [57], although events which draw attention to the project – press releases, broadcasts or major news – can drive large numbers of volunteers to the project for varying periods of time beyond this point [41]. Even so, our findings show that short-term projects gather significantly higher levels of both classifications and Talk comments during the 48 hour challenge period than other Zooniverse projects during their first 48 hours. Moreover, findings from the literature demonstrate that in 48 hours projects are capable of generating more classifications than are generated across six to nine months in other Zooniverse projects [39]. Similarly, in the case of a previous long-term project – Planet Four[11] – which used an initial 48 hour challenges, the challenge was responsible for the majority of classifications gathered even two years afterwards[12].

Naturally the short-term availability of these projects limits the overall amount of effort that individual volunteers can contribute and the total number of contributions that can be received [41, 57]. Nevertheless, short-term projects are associated with trade-offs in volunteer engagement. Cross-project effects, particularly between projects from similar domains have been noted as a key benefit of the Zooniverse, with new astrophysics projects benefiting on launch from an existing community of motivated individuals contributing to existing projects [8, 39]. In the astrophysics project Backyard Worlds: Planet Nine, our analysis suggests a significant reduction in task and discussion contributions during the 48 hour challenges associated with Planet Nine and Exoplanet Explorers. Nevertheless, this was not true of all – or even most – astrophysics projects during this period and most notably, the 48 hour periods naturally varied in terms of the day of the week which they covered. Moreover, participation in VCS projects is sporadic [50] and engagement in

[11] https://www.planetfour.org/

[12] http://blog.planetfour.org/2015/03/18/2-years-on-from-bbc-stargazing-live/

some projects was so low as to not be possible to analyse statistically. It is therefore quite possible that participation pulled from these less popular projects and ultimately more analysis is required to understand the extent to which new project launches – both short-term and more long-term – impact participation in other concurrent projects.

Nevertheless, there are also inevitable trade-offs in long-term volunteer engagement. Although volunteers contribute to VCS predominantly due to a desire to assist with science [51], the Snapshot Serengeti Zooniverse project reports that expected levels of redundancy are usually exceeded due to the large number of volunteers who classify images, leading to otherwise completed images being recirculated to ensure long-term engagement [65]. We find conflicting evidence of whether this would be beneficial within Planet Nine and Exoplanet Explorers[13]. While both projects attracted large numbers of volunteers, in Planet Nine, engagement within the project – both in task and Talk – sharply fell shortly after the completion of the 48 hour challenge event, as shown in Figures 3 and 4. Although it may seem inevitable that volunteers should cease engaging upon completing a project, this was not the case in Exoplanet Explorers, where classification levels initially fell but continued to rise over the subsequent 24 hours and where many users made their first Talk comments after the completion of the short-term challenge. Further analysis is required to identify whether short-term projects may curtail longer-term engagement, given the issues identified with Planet Four and Planet Nine.

While we found little evidence of specific negative impacts from short-term projects on volunteer engagement, there remains a significant skew in the distribution of effort expended by individual volunteers during the project period. Such skewed participation has been observed in a wide variety of other communities, such as Open Source software development [32], social networks [44], question and answer services [2] and collaborative wikis [60] to name but a few. Nevertheless, volunteer effort may not always be visible. More active participants – that is, those who contribute the most time and effort to initiatives – are more likely to be active in other elements of projects, including following mailing lists and news which may not lead to tangible results [15]. In turn, this inequality evolves throughout the life of a community, with community growth associated with a general increase in volunteer passivity and greater inequality of participation [46, 60]. The limited time available for contributions and growth in short-term projects, then, may have a beneficial role by limiting the development of a more extreme skew in volunteer participation levels. We suggest further research is needed to account for these potentially less visible forms of participation and consider how short-term projects may influence inequality of participation.

7.2 Impact on Discussion

Previous research has suggested that discussion interaction is not a significant factor in motivating participation within Zooniverse projects [51, 52]. Instead, discussion participation results predominantly from the need for feedback and information, leading to interactions that reinforce participants' desires to contribute to projects [10, 20]. The majority of Zooniverse participants do not make the move from task to Talk engagement and this is particularly true for newcomers to projects [21, 39]. Engagement with Talk, then, can be seen as synonymous with the meta-participation element of the motivational arc described by Crowston and Fagnot [9], which is only displayed by a small but dedicated number of longer-term volunteers in web-based collaborative systems. Indeed [63] suggest that volunteers – initially at least – did not understand the purpose of Talk.

Given these factors and the short-term nature of the analysed projects, it might be assumed that the time limitations placed on the project would have negative impacts on participants likelihood

[13]Note that while long-term engagement is by definition unnecessary in a short-term project, Exoplanet Explorers has since relaunched and Zooniverse projects rely on cross-project movement of volunteers [8]

to engage with Talk. Yet surprisingly, discussion activity within both Planet Nine and Exoplanet Explorers was significantly higher than in the initial 48 hours of other Zooniverse projects launched in the first 6 months of 2017, despite the limited time available and the emphasis on completing as many classifications as possible. Furthermore, a higher than average proportion of volunteers in Planet Nine and Exoplanet Explorers contributed to Talk, with 12.42% and 14.15% respectively compared with between 4.37% and 15.69% in the 10 Zooniverse projects analysed by Luczak-Roesch et al. [39]. Volunteers who contribute to Talk nevertheless contribute to projects for longer on average than those who contribute only to task, further suggesting a link between.

Why this might be is unclear. On the one hand, Jennett et al. [25] note that feedback is essential to allow volunteers to learn and in turn, encourage continued participation while Reeves et al. [53] note that feedback in virtual citizen science projects is predominantly delivered through Talk, not task elements. Yet in Planet Nine, volunteers were able to complete up to tens of thousands of classifications without the need to contribute to Talk and in both projects, task types were simple binary choices. These findings are similar to those of Jackson et al. [21], who found that a portion of volunteers choose to engage with Talk in the short-term, shortly after joining projects. We propose that further research is required to better understand the factors that influence this initial decision to engage with Talk and to explore the association between long-term engagement in VCS projects and the use of discussion.

7.3 Correlation Between Talk and Task

Our findings demonstrate a clear correlation between task and Talk contributions. In both projects, hourly Talk and task contribution rates are strongly correlated and in Planet Nine, after the final broadcast event Talk and task contribution levels each diminish similarly sharply in a similar time frame. In the case of Planet Nine then, at least, volunteers did not continue to use Talk beyond the short-term challenge period. On the one hand, this is not entirely surprising given the deliberate steps taken by the Zooniverse team to integrate *Talk* into task completion workflows [68]. Yet in Exoplanet Explorers, a significant proportion of volunteers decided to contribute to Talk outside of the short-term project period, as shown by Figure 6 and thus, Talk participation is not solely reliant on task.

While a positive correlation between task and Talk may be beneficial, there are significant potential trade-offs in short-term projects, with potentially adverse effects on the scientific elements of projects, particularly given the nature of replies and the relatively lengthy response times for individual subject comments. Should a comment not receive replies within the allotted time, then it is unlikely to receive a reply at all. A number of Zooniverse projects have played host to scientific discoveries which resulted not from task activity but from Talk discussions and Talk has previously been shown to be an important tool for volunteer retention and learning [6, 20]. While short-term projects also have positive impacts on Talk activity, care must be taken that they do not inhibit learning and feedback processes that rely on Talk activity.

7.4 Replies

Underlining the importance of replies, our findings suggest that receiving replies can have a number of beneficial outcomes in terms of participant behaviours. Task submissions levels and user lifespans are significantly higher for those users who received at least one reply to a comment than for those users who did not, while in the case of Exoplanet Explorers, users recieving a reply made more comments on average than those who did not. Analysis of Talk activity within Zooniverse projects suggests that only a minority of participants fill the role of starting and responding to discussion threads [21] and in both Exoplanet Explorers and Planet Nine, the majority of volunteers did not receive a reply to their Talk comments. While our results suggest that concerted efforts by

volunteers or project scientists to respond to and interact with Talk participants can encourage long-term engagement and higher levels of classification, Talk activity is sporadic and in the long term, volunteers are likely to leave projects, further reducing the size of the community who may respond to threads [57]. Although Zooniverse projects have moderator teams who can identify threads requiring scientist attention [68], the resources available to moderators are ultimately limited and average response rates exceed the average lifespan of users. In the short-term, however and particularly in these short-term projects, a short-term concerted effort by moderators, scientists and potentially volunteers to respond to threads can further boost contribution levels within projects and improve the efficiency and effectiveness of projects.

These findings are supported by observations made in other forms of online community participation. Within the online encyclopedia Wikipedia, new participants who were presented with welcoming messages and invited to engage in a volunteer forum known as the Wikipedia Teahouse contributed for significantly longer than new participants who did not. Similarly, a study of psychological wellbeing and user engagement in the user-generated open encyclopedia *Everything 2* found that feedback from other users – both criticism and validation – encouraged volunteers to continue participating, although a more important factor was a feeling of belonging to the community [73]. This feeling of belonging extends even beyond contributing and interacting users, as lurkers described feeling encouraged to participate after seeing the interactions of others [73]. This is an interesting area for further study.

7.5 Implications for Implementation

Currently, web-based citizen science projects designed to be available only in the short-term are relatively rare, although introducing short-term classification challenges into existing, established projects, such as 24 hour classification drives is somewhat common [53, 54]. Similarly, Bioblitzes are a form of citizen science which extends into both the online and offline sphere, where volunteers conduct field surveys aiming to record and identify the species present in a specific area within a set period of time [48]. The platform iNaturalist featured over 1000 Bioblitzes in 2017 alone, most notably the City Nature Challenge, which generates hundreds of thousands of submissions from around the world annually and features its own short-term project in the form of a five day classification identification challenge [14].

Our findings therefore have implications beyond the scope of short-term VCS initiatives, particularly given the importance of the opening days of citizen science projects and the similar issues outlined by Sauermann and Franzoni [57] surrounding sporadic and short-term volunteer engagement during project launches. For this reason and based on our analysis, we suggest that project scientists or representatives such as moderators should dedicate as much time as possible during the early days of a project to reading and responding to volunteer discussion messages. These efforts have a range of benefits for projects, including increasing volunteer contribution levels and securing long-term commitment from participants, but also a number of more qualitative benefits such as insight into potential problems within the project and the opportunity to build relationships and trust with participants which may pay off at later stages of the project [62].

Moreover, our results underline the importance of opportunities for meta-participation as described by Crowston and Fagnot [9]. While in the short-term participants can engage heavily with projects without such opportunities, meta-participation such as provided by Talk is associated with longer-term engagement of participants, even in short-term activities. It is particularly important to note that long-term participants complete the vast majority of work and are essential for ensuring

[14] see http://citynaturechallenge.org/

projects meet their goals [57]. Moreover, these discussion platforms play an important role in allowing participants to draw attention to potential discoveries, as in the case of Exoplanet Explorers. While these discoveries may be rare, they are also highly unpredictable [6]. Implementing these opportunities and platforms is therefore beneficial even for brief projects, in spite of potential difficulties associated with implementing such platforms[15].

The collaboration with Stargazing Live does not mark the first occasion on which broadcasting has been used to drive participation in citizen science. In the first days following the launch of Galaxy Zoo, there was little interest from volunteers until a radio interview with project scientists was broadcast, leading to a sharp and sudden but lasting increase in contributions and volunteer numbers [41]. The difficulties associated with recruitment through word of mouth have been well documented in online communities [18, 29]. Moreover, in citizen science participants recruited through word-of-mouth campaigns have been demonstrated to contribute fewer classifications and far less time than participants recruited through other methods. Similarly, the nature of these word-of-mouth messages has significant and potentially unexpected effects on participant engagement [36].

Our results suggest that the use of broadcast events can successfully engage a wide community of volunteers in citizen science projects within a short period of time, with broadcasts leading to temporary but highly significant bursts of contribution in both task and Talk, as shown by Figures 3 and 4. Locations in which broadcast events took place experience greater numbers of registering users and greater numbers of contributions, both in Task and in Talk. While it is not possible to know for certain what proportion of project contributors viewed – and subsequently decided to contribute due to – the live broadcasts, our findings suggest these played a critical role. In the case of the short-term projects, beyond launch emails, volunteers were predominantly engaged through the broadcast events[16] and we observe spikes in participation coinciding with the broadcast events.

Nevertheless, these bursts of behaviour should not be seen as unique to the context of citizen science, nor to broadcasting. Short periods of intense activity have been observed in a number of contexts, such as social media usage [16], online communication [47] and e-petitions [4]. In peer-production, the spread of news around disease outbreaks is associated with increased – but temporary – editing and contributing activity around articles within Wikipedia [1]. Pater et al. [49] particularly associate such bursty behaviours with individuals trying out new services or activities before losing interest and moving onto another. This in turn echoes the findings of Eveleigh et al. [12] regarding dabbling behaviours in citizen science and of Ponciano et al. [50] that many volunteers contribute for brief single sessions. Therefore, while broadcast events attract the attention of individual volunteers and raise awareness of the existence of projects, there is no evidence to suggest that they impact volunteer behaviours any more than other awareness campaigns.

7.6 Limitations

Given the relatively simple tasks described in the selected projects – and the relative rarity of images of interest – we have not considered the impact of time limits and short-term projects on contribution accuracy. This is, however, an interesting and important area for future research, particularly in the context of more difficult projects and tasks such as *EyeWire* in which it has been observed that volunteers take many hours and hundreds of contributions to learn to accurately contribute [27].

[15] such as the requirement for specialist knowledge and need for scientists' supervision or the efforts of moderators [68]

[16] While Zooniverse projects do make use of social media campaigns, these campaigns may be unsupervised, with volunteers choosing to share images and assets from projects they have enjoyed without encouragement. This makes observing and monitoring such campaigns difficult See, for example, [68].

In addition, our analysis relied on system log data, which did not allow for tracking anonymous traces associated with those users who choose not to register with the project/platform or to otherwise login, as described by Jackson et al. [22]. Although any anonymous users were responsible for the minority of classifications (7.89% in Exoplanet Explorers and 13.38% in Planet Nine) and anonymous users cannot contribute to Talk, an analysis of the behaviours associated with anonymous users in broadcasting and time-limited contexts is an important and interesting area for future research. Nevertheless, since anonymous users tend to contribute less heavily [22, 23], it can be assumed that the hypothesis testing results are generally accurate, although the effect sizes reported may be slightly lower than when considering anonymous users.

8 CONCLUSIONS

Our findings demonstrate that short-term citizen science projects can effectively and efficiently generate large volumes of volunteer submissions within a short period of time, without significantly sacrificing discussion activity. Moreover, our findings suggest that discussion and task completion are strongly correlated and engagement with discussion platforms is associated with higher levels of task contribution and longer user lifespans. With careful harnessing of Talk and task effort, we believe that short-term citizen science projects are an exciting and important area with the potential to overcome some of the commonly recurring issues that have effected web-based citizen science as a scientific methodology. Nevertheless, there are clear trade-offs to the use of such initiatives. The brief periods for which these projects are available naturally limit the total number of classifications received and size of the community engaged. Moreover, the short periods of engagement may curtail some of the benefits of long-term engagement, such as learning, meta-participation and development of stronger motivational factors such as altruism. At the same time, we have not accounted for the possibility of burn-out associated with increasing periods of heavy workloads. Before making wider use of short-term project cycles, project administrators must weigh these effects with potential gains in efficiency and with the likelihood and necessity for future engagement.

Moreover, we would emphasise the finding that the majority of volunteers who engaged with the short-term projects detailed within this paper did so to the exclusion of other projects, instead commonly only contributing to – at most – one other project. Further research is required to understand whether this effect is linked to the use of short-term crowdsourcing or is instead simply a result of the low levels of engagement that many participants display when engaging in citizen science activities.

One further area for exploration is the use of broadcasts to spread awareness and draw attention to citizen science projects. While our results suggest that broadcasts drive new users and contributions to projects, this effect appears to diminish with each additional broadcast. Nevertheless, these conclusions are complicated by the presence of cross-project effects and additional awareness campaigns. Additional research is required to understand how broadcasting drives participation and its effects on the short- and long-term engagement of volunteers, particularly outside of established partnerships such as with Stargazing Live.

Despite the specific context in which our analysis takes place, we also note observations that are applicable to VCS projects more broadly. In particular, we have shown that whether or not a volunteer receives engagement from the community and from project scientists in the form of replies to discussion comments significantly impacts the overall number of contributions that that individual makes within the project. Given the importance of the initial days of a project for its overall success, this discovery highlights the need for project scientists and the wider project community to actively engage with volunteers if scientists are to get the most out of volunteers.

ACKNOWLEDGMENTS

This research was supported by the ACTION project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 824603. We would like to thank Grant Miller and his colleagues at the Zooniverse for their help and support and the anonymous reviewers for their time and comments.

REFERENCES

[1] Reham Al Tamime, Richard Giordano, and Wendy Hall. 2018. Observing burstiness in Wikipedia articles during new disease outbreaks. In *Proceedings of the 10th ACM Conference on Web Science*. ACM, 117–126.
[2] Warren S Allen. 2018. Participation and knowledge exchange in a hybrid-economic software community. *Journal of the Knowledge Economy* 9, 3 (2018), 767–781.
[3] Aaron Bauer and Zoran Popović. 2017. Collaborative problem solving in an open-ended scientific discovery game. *Proceedings of the ACM on human-computer interaction* 1, CSCW (2017), 22.
[4] Lucas Böttcher, Olivia Woolley-Meza, and Dirk Brockmann. 2017. Temporal dynamics of online petitions. *PloS one* 12, 5 (2017), e0178062.
[5] Alex Bowyer, Chris Lintott, Greg Hines, Campbell Allan, and Ed Paget. 2015. Panoptes, a project building tool for citizen science. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing (HCOMPâĂŹ15). AAAI, San Diego, CA, USA*. 1–2.
[6] Carolin Cardamone, Kevin Schawinski, Marc Sarzi, Steven P Bamford, Nicola Bennert, C Megan Urry, Chris Lintott, William C Keel, John Parejko, Robert C Nichol, et al. 2009. Galaxy Zoo Green Peas: discovery of a class of compact extremely star-forming galaxies. *Monthly Notices of the Royal Astronomical Society* 399, 3 (2009), 1191–1205.
[7] Jessie L Christiansen, Ian JM Crossfield, Geert Barentsen, Chris J Lintott, Thomas Barclay, Brooke D Simmons, Erik Petigura, Joshua E Schlieder, Courtney D Dressing, Andrew Vanderburg, et al. 2018. The K2-138 System: A Near-resonant Chain of Five Sub-Neptune Planets Discovered by Citizen Scientists. *The Astronomical Journal* 155, 2 (2018), 57.
[8] Joe Cox, Eun Young Oh, Brooke Simmons, Chris Lintott, Karen Masters, Anita Greenhill, Gary Graham, and Kate Holmes. 2015. Defining and measuring success in online citizen science: A case study of Zooniverse projects. *Computing in Science & Engineering* 17, 4 (2015), 28–41.
[9] Kevin Crowston and Isabelle Fagnot. 2008. The motivational arc of massive virtual collaboration. In *Proceedings of the IFIP WG 9.5 Working Conference on Virtuality and Society: Massive Virtual Communities*. Citeseer, 1–2.
[10] Vickie Curtis. 2018. Exploring Online Citizen Science in Depth: A Tale of Three Projects. In *Online Citizen Science and the Widening of Academia*. Springer, 19–43.
[11] Alexis Dinno. 2015. Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *The Stata Journal* 15, 1 (2015), 292–300.
[12] Alexandra Eveleigh, Charlene Jennett, Ann Blandford, Philip Brohan, and Anna L Cox. 2014. Designing for dabblers and deterring drop-outs in citizen science. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2985–2994.
[13] Catherine O Fritz, Peter E Morris, and Jennifer J Richler. 2012. Effect size estimates: current use, calculations, and interpretation. *Journal of experimental psychology: General* 141, 1 (2012), 2.
[14] Anita Greenhill, Kate Holmes, Jamie Woodcock, Chris Lintott, Brooke D Simmons, Gary Graham, Joe Cox, Eun Young Oh, and Karen Masters. 2016. Playing with science: Exploring how game activity motivates users participation on an online citizen science platform. *Aslib Journal of Information Management* 68, 3 (2016), 306–325.
[15] Mordechai Muki Haklay. 2016. Why is participation inequality important? *European Handbook of Crowdsourced Geographic Information* (2016), 35.
[16] Jimpei Harada, David Darmon, Michelle Girvan, and William Rand. 2017. Prediction of elevated activity in online social media using aggregated and individualized models. In *Trends in Social Network Analysis*. Springer, 169–187.
[17] Ioanna Iacovides, Charlene Jennett, Cassandra Cornish-Trestrail, and Anna L Cox. 2013. Do games attract or sustain engagement in citizen science?: a study of volunteer motivations. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1101–1106.
[18] Alicia Iriberri and Gondy Leroy. 2009. A life-cycle perspective on online community success. *ACM Computing Surveys (CSUR)* 41, 2 (2009), 11.
[19] Corey Jackson, Kevin Crowston, Carsten Østerlund, and Mahboobeh Harandi. 2018. Folksonomies to support coordination and coordination of folksonomies. *Computer Supported Cooperative Work (CSCW)* 27, 3-6 (2018), 647–678.
[20] Corey Jackson, Carsten Østerlund, Kevin Crowston, Gabriel Mugar, and KD Hassman. 2014. Motivations for sustained participation in citizen science: Case studies on the role of talk. In *17th ACM Conference on Computer Supported*

Cooperative Work & Social Computing. Citeseer.
[21] Corey Jackson, Carsten Østerlund, Veronica Maidel, Kevin Crowston, and Gabriel Mugar. 2016. Which way did they go?: Newcomer movement through the Zooniverse. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 624–635.
[22] Corey Brian Jackson, Kevin Crowston, and Carsten Østerlund. 2018. Did they login?: Patterns of Anonymous Contributions in Online Communities. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 77.
[23] Caroline Jay, Robert Dunne, David Gelsthorpe, and Markel Vigo. 2016. To sign up, or not to sign up?: maximizing citizen science contribution rates through optional registration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 1827–1832.
[24] Charlene Jennett, Laure Kloetzer, Margaret Gold, and Anna L Cox. 2013. Sociability in Virtual Citizen Science. *on Designing and Evaluating Sociability in Online Video Games* (2013), 29.
[25] Charlene Jennett, Laure Kloetzer, Daniel Schneider, Ioanna Iacovides, Anna Cox, Margaret Gold, Brian Fuchs, Alexandra Eveleigh, Kathleen Methieu, Zoya Ajani, et al. 2016. Motivations, learning and creativity in online citizen science. *Journal of Science Communication* 15, 3 (2016).
[26] Firas Khatib, Seth Cooper, Michael D Tyka, Kefan Xu, Ilya Makedon, Zoran Popović, and David Baker. 2011. Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences* 108, 47 (2011), 18949–18953.
[27] Jinseop S Kim, Matthew J Greene, Aleksandar Zlateski, Kisuk Lee, Mark Richardson, Srinivas C Turaga, Michael Purcaro, Matthew Balkam, Amy Robinson, Bardia F Behabadi, et al. 2014. Space–time wiring specificity supports direction selectivity in the retina. *Nature* 509, 7500 (2014), 331.
[28] Margaret Kosmala, Andrea Wiggins, Alexandra Swanson, and Brooke Simmons. 2016. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment* 14, 10 (2016), 551–560.
[29] Robert E Kraut, Paul Resnick, Sara Kiesler, Moira Burke, Yan Chen, Niki Kittur, Joseph Konstan, Yuqing Ren, and John Riedl. 2012. *Building successful online communities: Evidence-based social design*. Mit Press.
[30] Marc J Kuchner, Jacqueline K Faherty, Adam C Schneider, Aaron M Meisner, Joseph C Filippazzo, Jonathan Gagné, Laura Trouille, Steven M Silverberg, Rosa Castro, Bob Fletcher, et al. 2017. The first brown dwarf discovered by the backyard worlds: Planet 9 citizen science project. *The Astrophysical Journal Letters* 841, 2 (2017), L19.
[31] Christopher Kullenberg and Dick Kasperowski. 2016. What is citizen science?–A scientometric meta-analysis. *PloS one* 11, 1 (2016), e0147152.
[32] Karim R Lakhani and Robert G Wolf. 2003. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. (2003).
[33] John Landers. 1981. Quantification in History, Topic 4: Hypothesis Testing II-Differing Central Tendency.
[34] Mirella Lapata. 2006. Automatic evaluation of information ordering: Kendall's tau. *Computational Linguistics* 32, 4 (2006), 471–484.
[35] Jeehyung Lee, Wipapat Kladwang, Minjae Lee, Daniel Cantu, Martin Azizyan, Hanjoo Kim, Alex Limpaecher, Snehal Gaikwad, Sungroh Yoon, Adrien Treuille, et al. 2014. RNA design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences* 111, 6 (2014), 2122–2127.
[36] Tae Kyoung Lee, Kevin Crowston, Mahboobeh Harandi, Carsten Østerlund, and Grant Miller. 2018. Appealing to different motivations in a message to recruit citizen scientists: results of a field experiment. *JCOM* 17, 01 (2018), A02–2.
[37] Chris J Lintott, Kevin Schawinski, William Keel, Hanny Van Arkel, Nicola Bennert, Edward Edmondson, Daniel Thomas, Daniel JB Smith, Peter D Herbert, Matt J Jarvis, et al. 2009. Galaxy Zoo:'Hanny's Voorwerp', a quasar light echo? *Monthly Notices of the Royal Astronomical Society* 399, 1 (2009), 129–140.
[38] Chris J Lintott, Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M Jordan Raddick, Robert C Nichol, Alex Szalay, Dan Andreescu, et al. 2008. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society* 389, 3 (2008), 1179–1189.
[39] Markus Luczak-Roesch, Ramine Tinati, Elena Simperl, Max Van Kleek, Nigel Shadbolt, and Robert Simpson. 2014. Why won't aliens talk to us? Content and community dynamics in online citizen science. In *Eighth International AAAI Conference on Weblogs and Social Media*.
[40] Thomas W MacFarland and Jan M Yates. 2016. Mann–whitney u test. In *Introduction to nonparametric statistics for the biological sciences using R*. Springer, 103–132.
[41] Alessandra Marli M Morais, Jordan Raddick, and Rafael D Coelho dos Santos. 2013. Visualization and characterization of users in a citizen science project. In *Next-Generation Analyst*, Vol. 8758. International Society for Optics and Photonics, 87580L.
[42] Gabriel Mugar, Carsten Østerlund, Katie DeVries Hassman, Kevin Crowston, and Corey Brian Jackson. 2014. Planet hunters and seafloor explorers: legitimate peripheral participation through practice proxies in online citizen science. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 109–119.
[43] Oded Nov, Ofer Arazy, and David Anderson. 2014. Scientists@ Home: what drives the quantity and quality of online citizen science participation? *PloS one* 9, 4 (2014), e90375.

[44] Oded Nov, Mor Naaman, and Chen Ye. 2010. Analysis of participation in an online photo-sharing community: A multidimensional perspective. *Journal of the American Society for Information Science and Technology* 61, 3 (2010), 555–566.

[45] Carsten Østerlund, Gabriel Mugar, Corey Brian Jackson, KD Hassman, and Kevin Crowston. 2014. Socializing the Crowd: Learning to talk in citizen science. In *Academy of Management Annual Meeting, OCIS Division, Philadelphia, PA*.

[46] Elliot Panek, Connor Hollenbach, Jinjie Yang, and Tyler Rhodes. 2017. Growth and inequality of participation in online communities: A longitudinal analysis. In *Proceedings of the 8th International Conference on Social Media & Society*. ACM, 51.

[47] Pietro Panzarasa and Moreno Bonaventura. 2015. Emergence of long-range correlations and bursty activity patterns in online communication. *Physical Review E* 92, 6 (2015), 062821.

[48] Sophie S Parker, Gregory B Pauly, James Moore, Naomi S Fraga, John J Knapp, Zachary Principe, Brian V Brown, John M Randall, Brian S Cohen, and Thomas A Wake. 2018. Adapting the bioblitz to meet conservation needs. *Conservation biology* 32, 5 (2018), 1007–1019.

[49] Jessica A Pater, Andrew D Miller, and Elizabeth D Mynatt. 2015. This Digital Life: A Neighborhood-Based Study of Adolescents' Lives Online. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2305–2314.

[50] Lesandro Ponciano, Francisco Brasileiro, Robert Simpson, and Arfon Smith. 2014. Volunteers' engagement in human computation for astronomy projects. *Computing in Science & Engineering* 16, 6 (2014), 52–59.

[51] M Jordan Raddick, Georgia Bracey, Pamela L Gay, Chris J Lintott, Carie Cardamone, Phil Murray, Kevin Schawinski, Alexander S Szalay, and Jan Vandenberg. [n.d.]. Galaxy Zoo: Motivations of Citizen Scientists. *Astronomy Education Review* 12, 1 ([n. d.]).

[52] Jason Reed, M Jordan Raddick, Andrea Lardner, and Karen Carney. 2013. An exploratory factor analysis of motivations for participating in Zooniverse, a collection of virtual citizen science projects. In *2013 46th Hawaii International Conference on System Sciences*. IEEE, 610–619.

[53] Neal Reeves, Ramine Tinati, Sergej Zerr, Max G Van Kleek, and Elena Simperl. 2017. From crowd to community: a survey of online community features in citizen science projects. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 2137–2152.

[54] Neal Reeves, Peter West, and Elena Simperl. 2018. A game without competition is hardly a game": The impact of competitions on player activity in a human computation game. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*.

[55] Frauke Rohden, Christopher Kullenberg, Niclas Hagen, and Dick Kasperowski. 2019. Tagging, Pinging and Linking–User Roles in Virtual Citizen Science Forums. *Citizen Science: Theory and Practice* 4, 1 (2019).

[56] Dana Rotman, Jenny Preece, Jen Hammock, Kezee Procita, Derek Hansen, Cynthia Parr, Darcy Lewis, and David Jacobs. 2012. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. ACM, 217–226.

[57] Henry Sauermann and Chiara Franzoni. 2015. Crowd science user contribution patterns and their implications. *Proceedings of the national academy of sciences* 112, 3 (2015), 679–684.

[58] Shlomo S Sawilowsky. 2009. New effect size rules of thumb. (2009).

[59] Megan E Schwamb, Jerome A Orosz, Joshua A Carter, William F Welsh, Debra A Fischer, Guillermo Torres, Andrew W Howard, Justin R Crepp, William C Keel, Chris J Lintott, et al. 2013. Planet hunters: A transiting circumbinary planet in a quadruple star system. *The Astrophysical Journal* 768, 2 (2013), 127.

[60] Abel Serrano, Javier Arroyo, and Samer Hassan. 2018. Participation Inequality in Wikis: A Temporal Analysis Using WikiChron.. In *OpenSym*. 12–1.

[61] Sally Shuttleworth. 2017. Life in the Zooniverse: Working with Citizen Science. *Journal of Literature and Science* 10, 1 (2017).

[62] Elena Simperl, Neal Reeves, Chris Phethean, Todd Lynes, and Ramine Tinati. 2018. Is virtual citizen science a game? *ACM Transactions on Social Computing* 1, 2 (2018), 6.

[63] Artemis Skarlatidou, Alexandra Hamilton, Michalis Vitos, and Muki Haklay. 2019. What do volunteers want from citizen science technologies? A systematic literature review and best practice guidelines. *JCOM: Journal of Science Communication* 18, 1 (2019).

[64] James Sprinks, Jessica Wardlaw, Robert Houghton, Steven Bamford, and Jeremy Morley. 2017. Task Workflow Design and its impact on performance and volunteers' subjective preference in Virtual Citizen Science. *International Journal of Human-Computer Studies* 104 (2017), 50–63.

[65] Alexandra Swanson, Margaret Kosmala, Chris Lintott, Robert Simpson, Arfon Smith, and Craig Packer. 2015. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific data* 2 (2015), 150026.

[66] Ramine Tinati, Markus Luczak-Roesch, Elena Simperl, Nigel Shadbolt, and Wendy Hall. 2015. '/Command'and Conquer: Analysing Discussion in a Citizen Science Game. In *Proceedings of the ACM Web Science Conference*. ACM, 26.
[67] Ramine Tinati, Elena Simperl, Markus Luczak-Rösch, Max Van Kleek, and Nigel Shadbolt. 2014. Collective intelligence in citizen science–a study of performers and talkers. *Proceedings of Collective Intelligence 2014* (2014).
[68] Ramine Tinati, Max Van Kleek, Elena Simperl, Markus Luczak-Rösch, Robert Simpson, and Nigel Shadbolt. 2015. Designing for citizen data analysis: a cross-sectional case study of a multi-domain citizen science platform. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 4069–4078.
[69] Laura Trouille. 2016. From Clicks to Publications: How the Public is Changing the Way We Do Research. In *Frank N. Bash Symposium 2015*, Vol. 261. SISSA Medialab, 009.
[70] David Watson and Luciano Floridi. 2018. Crowdsourced science: sociotechnical epistemology in the e-research paradigm. *Synthese* 195, 2 (2018), 741–764.
[71] Andrea Wiggins and Kevin Crowston. 2011. From conservation to crowdsourcing: A typology of citizen science. In *2011 44th Hawaii international conference on System Sciences*. IEEE, 1–10.
[72] Andrea Wiggins and Kevin Crowston. 2015. Surveying the citizen science landscape. *First Monday* 20, 1 (2015).
[73] Donghee Yvette Wohn and Cliff Lampe. 2018. Psychological Wellbeing as an Explanation of User Engagement in the Lifecycle of Online Community Participation. In *Proceedings of the 2018 ACM Conference on Supporting Groupwork*. ACM, 184–195.
[74] James Wynn. 2017. *Citizen science in the digital age: rhetoric, science, and public engagement*. University of Alabama Press.
[75] Weichao Xu, Yunhe Hou, YS Hung, and Yuexian Zou. 2013. A comparative analysis of Spearman's rho and Kendall's tau in normal and contaminated normal models. *Signal Processing* 93, 1 (2013), 261–276.

Received April 2019; revised June 2019; accepted August 2019.

A APPENDIX A - FIGURES

Table 6. Kruskal Wallis H test outcomes comparing hourly contribution levels before, during and after the launch of Planet Nine and Exoplanet Explorers for classifications and Talk.

Name	Period	H - Classifications	p - Classifications	H - Talk Comments	p - Talk Comments
Astronomy Rewind	P9	12.77	0.002	5.47	0.065
	EE	1.85	0.398	3.32	0.190
Backyard Worlds: Planet Nine	P9	46.89	<0.001	8.79	0.012
	EE	37.84	<0.001	21.05	<0.001
Camera Catalogue	P9	5.78	0.820	2.11	0.348
	EE	0.40	0.055	6.12	0.047
Chicago Wildlife Watch	P9	4.22	0.121	18.92	<0.001
	EE	4.50	0.105	7.38	0.025
Cleveland Metroparks	P9	2.33	0.311	1.72	0.421
	EE	33.44	<0.001	3.41	0.181
Comet Hunters	P9	15.90	<0.001	1.53	0.466
	EE	1.17	0.557	4.07	0.131
Gravity Spy	P9	19.49	<0.001	0.04	0.979
	EE	2.25	0.325	7.15	0.028
Milky Way Project	P9	4.04	0.132	8.01	0.018
	EE	0.426	0.808	6.92	0.031
Muon Hunters	P9	6.82	0.033	8.29	0.016
	EE	9.74	0.008	3.85	0.146
Steller Watch	P9	9.78	0.008	0.08	0.961
	EE	3.10	0.212	2.59	0.274

B APPENDIX B

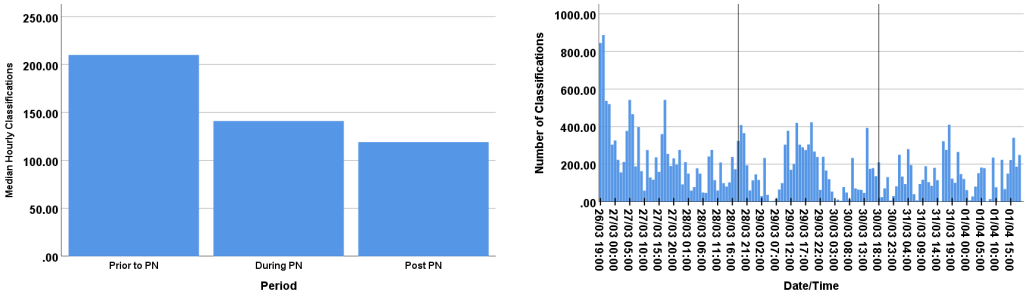

Fig. 7. Median and individual hourly classification counts for the Astronomy Rewind project for the 48 hours prior to, during and after the launch of the Planet Nine classification challenge. Average classification levels are statistically significantly higher prior to PN when compared with during (z=-2.57, p=0.03, d=0.44) and after the classification challenge (z=-3.43, p=0.002, d=0.60).

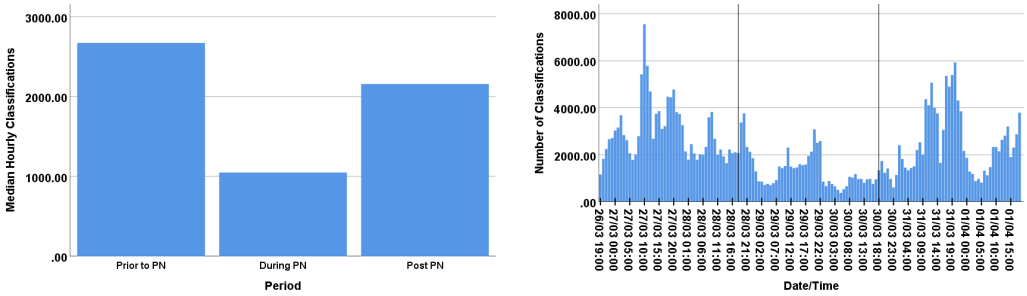

Fig. 8. Bar charts showing median and individual hourly classification counts for the Backyard Worlds: Planet Nine project for the 48 hours prior to, during and after the launch of the Planet Nine classification challenge. Average classification levels are statistically significantly lower during PN when compared with the 48 hours prior to (z=-6.72, p<0.001, d=1.35) and after the classification challenge (z=-4.50, p<0.001, d=0.81).

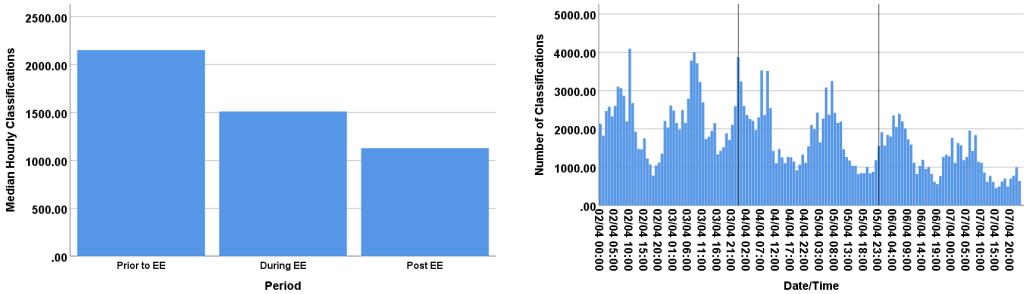

Fig. 9. Bar charts showing median and individual hourly classification counts for the Backyard Worlds: Planet Nine project for the 48 hours prior to, during and after the launch of the Exoplanet Explorers classification challenge. Average classification levels are statistically significantly higher prior to EE when compared with during (z=-2.73, p=0.019, d=0.47) and after the classification challenge (z=-6.14, p<0.001, d=1.19).

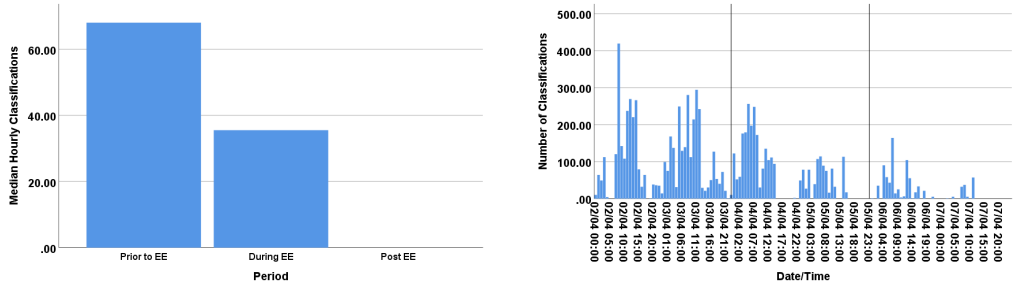

Fig. 10. Bar charts showing median and individual hourly classification counts for the Cleveland Metroparks project for the 48 hours prior to, during and after the launch of the Exoplanet Explorers classification challenge. Average classification levels are statistically significantly lower post EE when compared with prior to (z=-3.10, p=0.006, d=0.53) and during the classification challenge (z=-3.02, p=0.008, d=0.52).

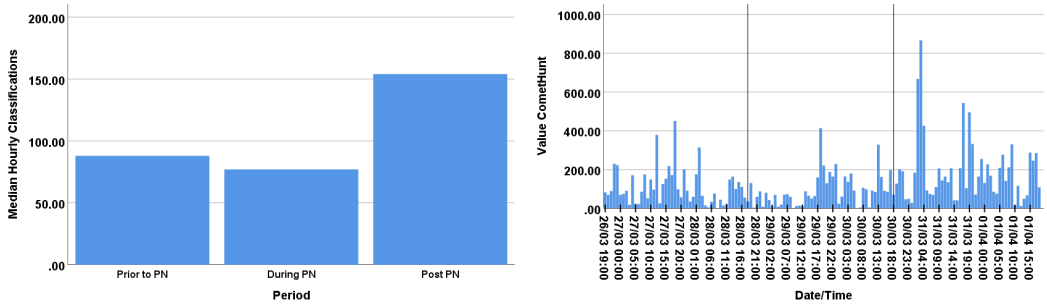

Fig. 11. Bar charts showing median and individual hourly classification counts for the Comet Hunters project for the 48 hours prior to, during and after the launch of the Planet Nine classification challenge. Average classification levels are statistically significantly higher post PN when compared with prior to (z=-2.89, p=0.012, d=0.50) and during the classification challenge (z=-3.82, p<0.001, d=0.67).

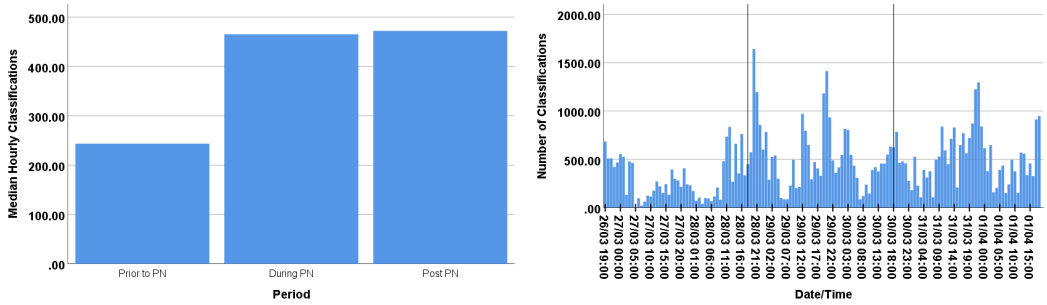

Fig. 12. Bar charts showing median and individual hourly classification counts for the Gravity Spy project for the 48 hours prior to, during and after the launch of the Planet Nine classification challenge. Average classification levels are statistically lower prior to PN when compared with during (z=-3.62, p=0.001, d=0.63) and after the classification challenge (z=-4.00, p<0.001, d=0.71).

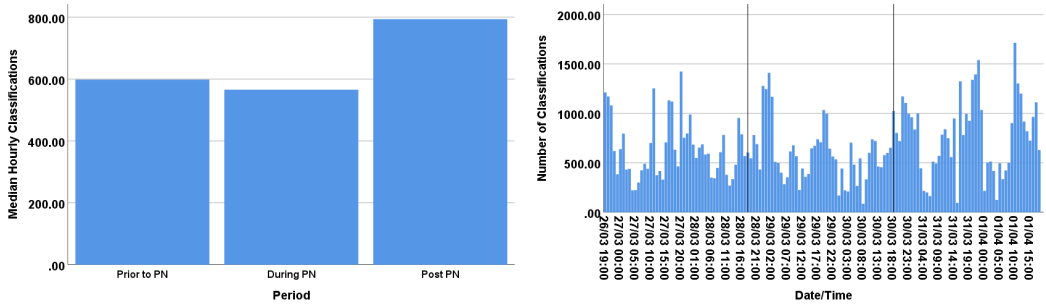

Fig. 13. Bar charts showing median and individual hourly classification counts for the Muon Hunters project for the 48 hours prior to, during and after the launch of the Planet Nine classification challenge. Average classification levels are statistically higher during PN when compared with after the classification challenge (z=-2.48, p=0.040, d=0.42).

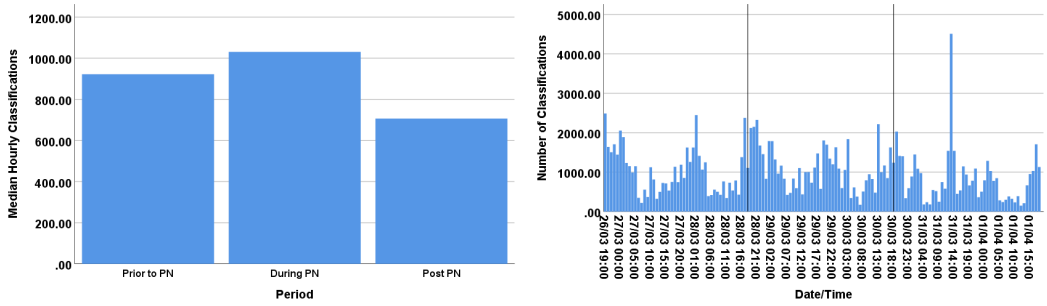

Fig. 14. Bar charts showing median and individual hourly classification counts for the Steller Watch project for the 48 hours prior to, during and after the launch of the Planet Nine classification challenge. Average classification levels are statistically lower post PN when compared with during the classification challenge (z=-3.05, p=0.007, d=0.53).

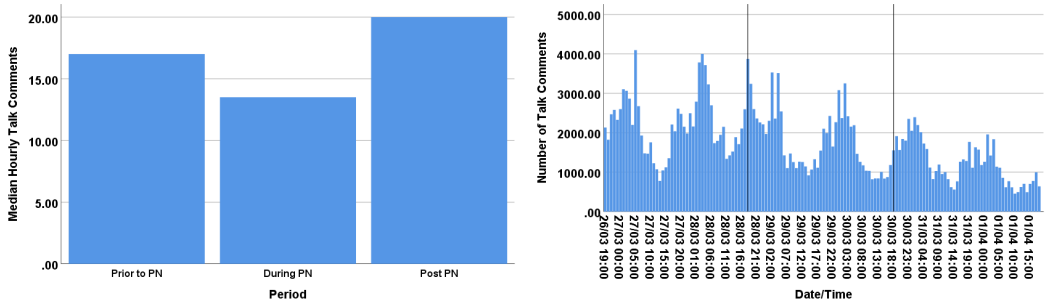

Fig. 15. Bar charts showing median and individual hourly Talk comment counts for the Backyard Worlds: Planet Nine project for the 48 hours prior to, during and after the launch of the Planet Nine classification challenge. Average classification levels are statistically lower during PN when compared with after the classification challenge (Z=-1.23, p=0.009, d=0.21).

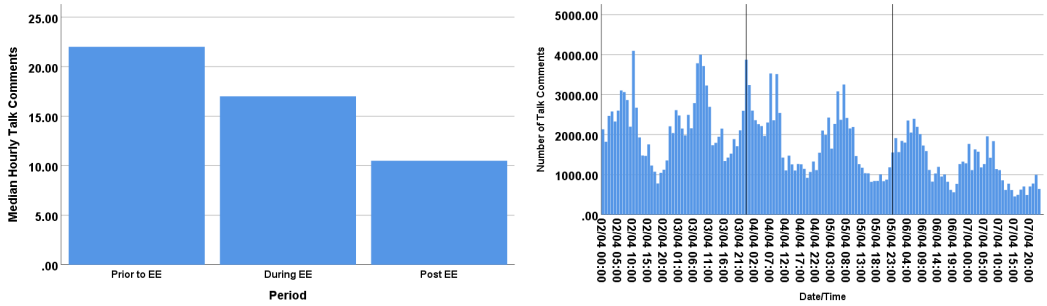

Fig. 16. Bar charts showing median and individual hourly Talk comment counts for the Backyard Worlds: Planet Nine project for the 48 hours prior to, during and after the launch of the Exoplanet Explorers classification challenge. Average classification levels are statistically lower after EE when compared with the period prior to (Z=-4.40, p<0.001, d=0.79) and during the classification challenge (Z=-3.32, p=0.003, d=0.58).

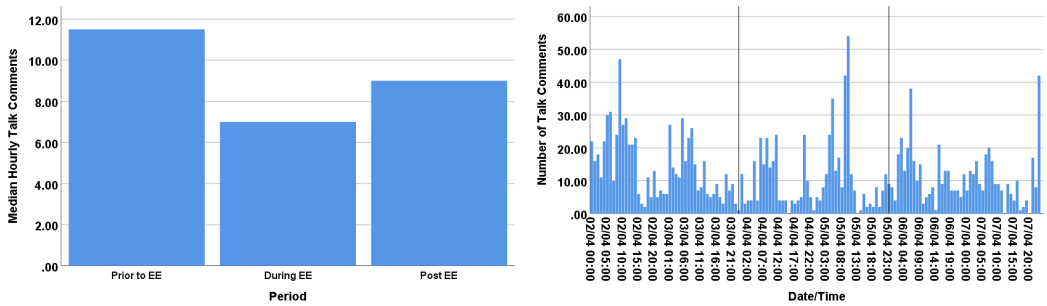

Fig. 17. Bar charts showing median and individual hourly Talk comment counts for the Camera Catalogue project for the 48 hours prior to, during and after the launch of the Exoplanet Explorers classification challenge. Average classification levels are statistically lower during EE when compared with the period prior to the classification challenge (Z=-2.46, p=0.042, d=0.42).

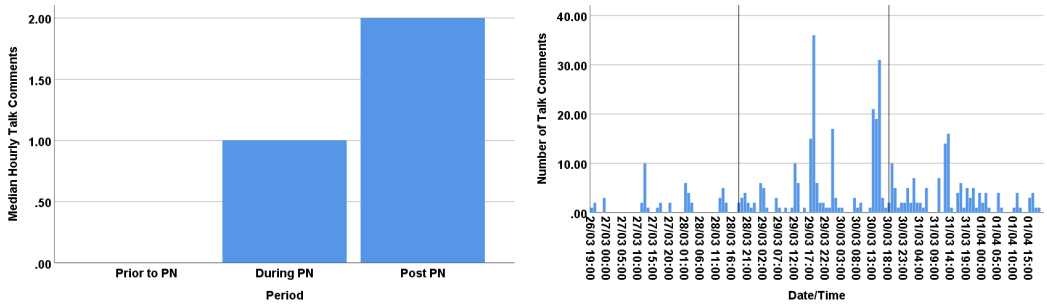

Fig. 18. Bar charts showing median and individual hourly Talk comment counts for the Chicago Wildlife Watch project for the 48 hours prior to, during and after the launch of the Planet Nine classification challenge. Average classification levels are statistically higher prior to PN when compared with the period during (Z=-3.71, p=0.001, d=0.65) and after the classification challenge (Z=-3.82, p<0.001, d=0.67).

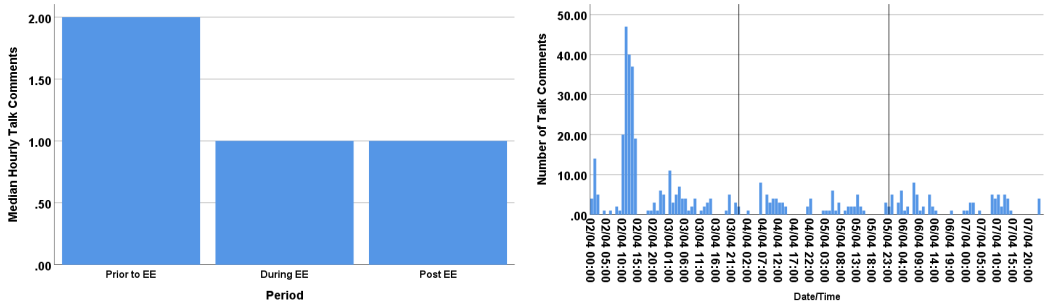

Fig. 19. Bar charts showing median and individual hourly Talk comment counts for the Chicago Wildlife Watch project for the 48 hours prior to, during and after the launch of the Exoplanet Explorers classification challenge. Average classification levels are statistically higher during EE when compared with the period during which the classification challenge took place (Z=-2.47, p=0.041, d=0.42).

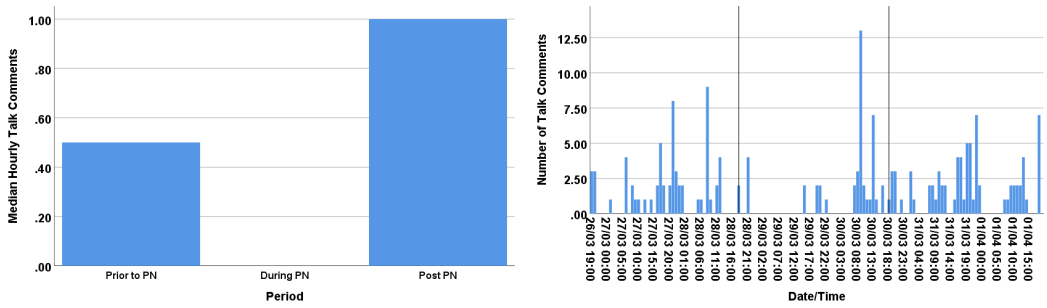

Fig. 20. Bar charts showing median and individual hourly Talk comment counts for the Milky Way Project for the 48 hours prior to, during and after the launch of the Planet Nine classification challenge. Average classification levels are statistically significantly lower during PN when compared with the period after which the classification challenge took place (Z=-2.80, p=0.015, d=0.48).

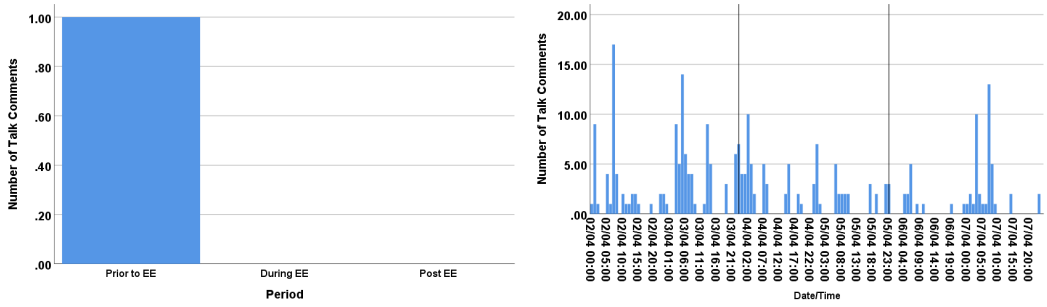

Fig. 21. Bar charts showing median and individual hourly Talk comment counts for the Milky Way Project for the 48 hours prior to, during and after the launch of the Exoplanet Explorers classification challenge. Average classification levels are statistically significantly lower after EE when compared with the period prior to the classification challenge (Z=-2.63, p=0.026, d=0.45).

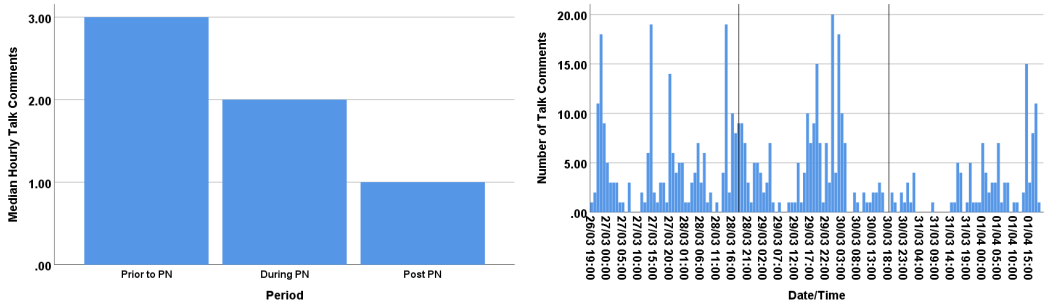

Fig. 22. Bar charts showing median and individual hourly Talk comment counts for the Steller Watch project for the 48 hours prior to, during and after the launch of the Planet Nine classification challenge. Average classification levels are statistically significantly higher prior to PN when compared with the period after which the classification challenge took place (Z=-2.72, p=0.020, d=0.47).