"A Game Without Competition is Hardly a Game": The Impact of Competitions on Player Activity in a Human Computation Game

Neal Reeves and **Peter West** and **Elena Simperl**
Web and Internet Science, Electronics and Computer Science, University of Southampton

Abstract

Virtual citizen science (VCS) projects enable new forms of scientific research using crowdsourcing and human computation to gather and analyse large-scale datasets. To attract and sustain the number of participants and levels of participation necessary to achieve research aims, some VCS projects have introduced game elements such as competitions to tasks. However, we still know very little about how some game elements, particularly competitions, influence participation rates. To investigate the impact of game elements on player engagement, we conducted a two-part mixed-methods study of EyeWire, a VCS game. First, we interviewed EyeWire designers to understand their rationale for introducing competitions. Guided by their answers, we analysed two datasets of EyeWire user task contributions and chat logs to assess the effectiveness of competitions in achieving designers' goals. Our findings contribute to the growing understanding of how competitions influence participant activity in human computation initiatives and socio-technical systems such as VCS.

Introduction

Virtual citizen science (VCS) projects are a form of human computation, which draws on large-scale crowdsourcing methodologies to gather and analyse datasets for scientific research purposes (Wiggins and Crowston 2011). As modern technology has enabled and demanded the collection of large volumes of data and images, traditional scientific workflows where small teams of scientists collect and analyse data have becoming increasingly insufficient (Lintott et al. 2008). The use of citizen science methodologies has been increasingly suggested as a solution to such issues, with participants preparing datasets or training algorithms for future research use (Khatib et al. 2011; Lintott et al. 2008). Relying on the 'wisdom of crowds', VCS initiatives serve as socio-technical systems, which draw on a combination of human and machine agents to generate and process data (Jennings et al. 2014; Michelucci 2013). Furthermore, projects serve as online communities, where participants may use computer-mediated communication platforms to interact with one another and further engage with scientific research (Luczak-Roesch et al. 2014; Østerlund et al. 2014; Oliveira, Jun, and Reinecke 2017).

However, as with other human computation initiatives, designing suitable VCS projects remains a pervasive research topic. Projects must balance a number of trade-offs to generate sufficiently accurate data, while attracting and supporting large communities of participants in undertaking unfamiliar and potentially complex tasks (Tinati et al. 2015b). Datasets may consist of hundreds of thousands to millions of artefacts, each potentially requiring analysis by multiple participants to ensure the validity of responses (Tinati et al. 2015b; Wiggins et al. 2011). As with many other web-based collaborative communities, a minority of participants often produce the majority of contributions, with a large proportion of users contributing for single, brief sessions (Ponciano et al. 2014; Sauermann and Franzoni 2015). To address problems, an increasing number of projects are turning to gamification and game elements as a means to attract and sustain volunteer engagement with VCS and increase levels of participation (Bowser et al. 2013; Curtis 2014; Reeves et al. 2017).

Yet the use of gamification for human computation is not without its challenges. The effectiveness of game elements depends on the context of platforms and the personality characteristics of participants (Jia et al. 2016; 2017). Gamified elements often introduce extrinsic reward mechanisms such as points, badges and physical prizes (Reeves et al. 2017; Seaborn and Fels 2015). However, studies have shown that participants have different motivations; while some align well with these elements (Jia et al. 2016), other volunteers' intrinsic motivations are negatively impacted by the use of extrinsic factors (Greenhill et al. 2016). This is problematic in VCS, where projects predominantly rely on volunteers' intrinsic motivations to encourage and motivate contributions (Cox et al. 2015; Rotman et al. 2012). In particular, we know less about the effects of competitive game elements such as global leaderboards and contests. Introducing competitive elements has been observed to make tasks stressful and demotivating, as individual players gain unassailable leads over fellow participants (Eveleigh et al. 2013). The impact of point-based game design and the use of interplayer competitions has been suggested to be minimal, in contrast to narrative and framing devices (Lieberoth 2015; Mekler et al. 2013a; Prestopnik and Tang 2015). Specific to VCS, game elements risk presenting simplistic ideas of scientific research (Sandbrook, Adams, and Monteferri 2015), negatively influencing engagement in VCS (Eveleigh et al. 2013; Ponti, Hillman, and Stankovic 2015).

In this paper, we explore the impact of competition events in *EyeWire*[1], a Virtual Citizen Science Game With a Purpose. We first conduct a series of structured reflection interviews with six key members of the EyeWire project team, in order to identify the factors motivating the implementation of competition events within the project. Using the results of this interview process as an analytical framework, we then assess the effectiveness of competitions held within a one year period compared with non-competitive periods, across factors such as number of contributions, number of participants, performance of additional tasks and social interaction. Our work builds on previous research in this area by drawing on the experiences of VCS designers and offering quantitative evidence of the impacts of competitions on player activity levels in VCS.

Background and Related Work

Citizen Science

Citizen science (CS) describes the engagement of volunteer participants in the scientific research process (Dickinson, Zuckerberg, and Bonter 2010). Volunteers undertake several activities, including data collection, classification and analysis (Tinati et al. 2015b). In some forms of citizen science, participants may even co-design experiments and research questions, drawing conclusions and publishing findings (Haklay 2013). As access to the Web has grown, so has a new form of citizen science: *Virtual Citizen Science*. In this form, rather than gathering data and making records, participants contribute their knowledge through crowd-sourced human computation activities, delivered through web-based and mobile portals (Wiggins and Crowston 2011).

As a human computation process, VCS is characterised by a number of tradeoffs between factors such as engagement, ease of use and data quality (Aoki et al. 2017; Sprinks et al. 2015b). For example, while VCS introduces new concepts and tasks, completion rates of tutorials are low, resulting in trade-offs or the need to support human computation through algorithms and software (Mattos et al. 2014). Findings from a survey of participant motivations in the Zooniverse platform carried out by Sprinks et al suggest that players require feedback to resolve task performance concerns and to feel valued (Sprinks et al. 2015a).

VCS is also highly dependent on the suitability of tasks provided to volunteers. Breaking overarching macrotasks into smaller, repeateable microtask actions is a distinct challenge - one that has been shown to increase the overall time taken to complete tasks, yet lead to an increase in the quality of outputs and user experience (Cheng et al. 2015). Design decisions must also consider a wide variety of factors and draw on a number of skillsets and array of disciplinary knowledge. Tinati et al. synthesised design recommendations from interview sessions with key members of the design team responsible for the Zooniverse platform (Tinati et al. 2015b). These recommendations show the interplay of complex and potentially unpredictable factors in motivating VCS engagement. Beyond engagement, VCS projects are also subject to a variety of data quality and accuracy concerns. Participants can nonetheless be a valuable resource for ensuring the validity of responses, but this increases the workload expected of the community and thus requires further engagement. For example, repetition of individual microtasks by multiple users is a common form of data validation in VCS (Wiggins et al. 2011).

Gamification

Gamification describes the use of "game elements in non-gaming contexts" (Deterding et al. 2011). We make the distinction between gamification and the similar, yet distinct, *Games With A Purpose*, where participants carry out human computation tasks as part of a casual game experience. (Siu, Zook, and Riedl 2014). While human computation initiatives such as Citizen Science projects may make use of gamified elements (such as points and rewards) to encourage player engagement with tasks, it is the diegetic nature of the elements in Games With A Purpose which sets them apart from other project varieties (Prestopnik, Crowston, and Wang 2014). Gamification has been employed in a wide array of contexts, from education and research to business and marketing (Seaborn and Fels 2015) and a significant focus of gamification research concerns evaluating the impact of diverse game elements in varying contexts and platforms (Jia et al. 2017; Kumar 2013; Mekler et al. 2013b).

A diverse set of studies have considered the role that gamification and games play in motivating participation in a number of human computation games (Bowser et al. 2013; Eveleigh et al. 2013; Iacovides et al. 2013). Results from such findings have been mixed, with players identifying both positive and negative aspects of gamification in VCS. However, these studies draw on survey-based methodologies and predominantly focus on qualitative measures of volunteer participation, while quantitative analyses have suggested that self-reported motivations do not align with volunteer contributions (Mekler et al. 2013a). Our paper aims to contribute to this area of research by providing quantiative analysis of the impact of competitions on particpation.

We further note a number of studies and findings unrelated to VCS within the wider literature of relevance to our analysis, which we seek to build on to further the understanding of gamification on engagement and in HCI more broadly. Jia et al. explored the impact of individual gamification elements on participants according to personality characteristics (Jia et al. 2016). In team-based task completion, partnering with highly effective teammates has been found to reduce players own intrinsic motivations and perceptions of tasks (Luu and Narayan 2017). Brouwer conducted a survey of professional work teams in the Netherlands around the subject of intra-team competition and found that such competition has both negative and positive effects on team performance, through increased task complexity but reduced psychological safety (Brouwer 2016).

More specific to Games With A Purpose, but in a non-citizen science context, Siu et al. explored the impact of

[1]https://EyeWire.org/explore

competition-based scoring and reward mechanisms compared with collaboration-based reward mechanisms (Siu, Zook, and Riedl 2014). In particular, the authors found that intrinsic motivation and associated contest-design decisions were key to driving participation in competitions, but nevertheless must be balanced with extrinsic factors. In a similar domain to our research, Zheng et al. analysed motivations for participating in crowdsourcing competitions in paid microtask work, showing that engagement is linked to the nature of contest demands (Zheng, Li, and Hou 2011). Similarly, introducing cognitively demanding challenges to games has been found to increase player activity, while physically challenging tasks do not (Cox et al. 2012). Our work seeks to build on these studies by exploring the impact of adding temporary competitions to a specifically VCS context, where both altruistic and intrinsic motivations drive participation.

EyeWire

EyeWire is a crowdsourced citizen science Game With A Purpose, where participants contribute to identifying the presence of neuron cells in Magnetic Resonance images of mouse retina, mapping each image for further research (Kim et al. 2014; Tinati et al. 2017). Players do so by ‘tracing’ a three-dimensional section of an imaged neuron cell by manipulating two-dimensional cross-sectional images (see Figure 1). These 3D segments are known as cubes and it is these cubes that players submit to the project. Each cube starts with a ‘seed’ generated algorithmically and players must trace the path of the neuron from this seed through the cube. Multiple players trace each cube to generate an overall consensus as a means to improve the accuracy of submissions (Kim et al. 2014).

In order to convert this relatively repetitive task into an engaging game, EyeWire draws on a number of features. Players earn points for each cube traced, changing position within an integrated leaderboard, displaying daily, weekly and monthly points earned by each contributing user (Reeves et al. 2017; Tinati et al. 2017). The EyeWire science team regularly organises competitions, where players contribute through the main EyeWire interface as a team or as individuals (Tinati et al. 2017). Less frequently, the team organises large scale events, where a number of competition categories occur within a week-long period, united by a common storyline[2].

As well as the scientific task, EyeWire features a variety of discussion platforms including an integrated instant messenger chat service, forum and Wiki. Studies of chat participation have revealed that the instant messenger chat is vital for knowledge sharing, allowing new participants to learn from experienced players, promoting collaboration on tasks and encouraging players to continue to participate in EyeWire (Tinati, Simperl, and Luczak-Roesch 2017) and the chat interface is heavily used by a small but very active proportion of players (Tinati et al. 2015a).

The EyeWire task is cognitively complex and demands a significant time investment from participants to master the game. Rather than simply gathering or analysing data, volunteers in EyeWire must solve problems: identifying unpredictable branching neuron pathways, which span thousands of cubes (Lieberoth et al. 2014; Kim et al. 2014). EyeWire does not require participants to possess domain expertise or previous training (Borrett and Hughes 2016; Tinati et al. 2015a) but barriers for successful participation in the project are relatively high as players must first develop relevant skills (Lieberoth et al. 2014). Although the game provides a short compulsory tutorial, the most successful and accurate contributors have invested “tens of hours” over the course of “thousands of cubes” (Kim et al. 2014).

[2] see: http://blog.eyewire.org/competitions-the-master-guide/

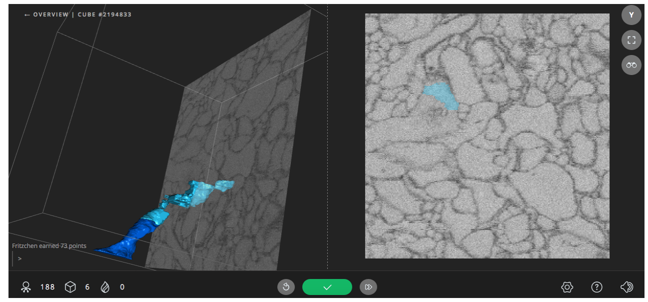

Figure 1: EyeWire tracing interface, showing left hand 3D cube view with seed and tracing, right hand 2D cross-sectional view and bottom left corner live chat overlay.

Data and Methods

To begin our analysis, we first sought to understand the role of competitions in the EyeWire project, as well as the needs of the project scientists and conditions for project success. To this end, we conducted a series of interviews with six members of the EyeWire team, over the course of two days. These participants included the project director, three ‘game masters’ (responsible for the day to day management of the community and monitoring manual and automated game processes), one designer and one developer. The six individuals represent the whole of the full-time EyeWire team and offered a diverse range of specialist knowledge about the history and outcomes of the project.

Following an approach used in a similar study (Tinati et al. 2015b), interviews used a structured reflection approach. During each session, participants were encouraged to share views, observations and experiences from the lifetime of the project. To guide participants and to provide a starting point for deeper discussions, we provided a series of questions framed around three key themes, including design decisions resulting in the use of competitions, expected and desired behaviours and measures and metrics of project success. After transcribing each interview, two researchers individually coded each of the transcripts, using our initial framework of three themes. After this process was complete, we compared the coded transcripts and identified and corrected areas of disagreement. From these coded transcripts, we derived a set of hypothesised participation patterns based on the experiences of the project team, which we identify as H.1 - H.6 (see *Interview Results - Desired Outcomes*).

Using the hypotheses outlined through the structured reflection process, we collected datasets from the EyeWire platform for quantitative analyses aimed at confirming the validity of each hypothesis. An initial dataset identified each player who contributed to Eyewire, as well as their number of contributions (in terms of points and cubes) for each 24 hour period from August 2016 - August 2017, including a separate log of final scores for each competition during this period. In response to the chat-related hypotheses put forward by the team, we also gathered a second dataset over a (91 day) period from June - August 2017. This dataset consisted of a time-stamped datastream containing discussion messages, achievements and system messages for all users active during these 91 days. Since players can contribute to the *Trivia* class through chat commands and the use of a bot set up for this competition class, we were unable to completely access data for this competition class. As a result and given the lack of hypotheses regarding this competition class, we removed all visible trivia entries from the chat data before further analysis.

For each dataset, we calculated relevant descriptive statistics, including mean total player numbers, cube contribution numbers, chat participant and chat message counts. Since competition participation in EyeWire is entirely voluntary, we distinguish between *competitors* who choose to participate in a given competition and *non-competitors* who do not. For each hypothesis (H1-H6) we defined suitable comparative test conditions. Using these conditions, we selected all available data from the relevant dataset - i.e, for *H1*, using dataset one we selected player numbers for all days on which competitions were held and compared these with all days on which they were not. Given the non-normal distributions and unpaired nature of the datasets involved, we used the non-parametric Mann-Whitney U test to evaluate each hypothesis. We report these test statistics using the Z-test, for ease of interpreation. A further set of additional tests (A1 - A4) were carried out to explore the impact of competitions on non-competing players. In addition to test statistics, we calculated a measure of effect size, Cohen's d, for each hypothesis, which we interpret using Sawilowsky's revisions of Cohen's original recommendations (Sawilowsky 2009). A summary of tests carried out, statistics and findings can be seen in table 3.

Results - Interviews

Design Requirements

The most significant factor influencing the implementation of regular competitions in the EyeWire project was to overcome difficulties arising from the nature of VCS games. During the interviews, the team described a range of characteristics they associated with successful and engaging games, which they had struggled to introduce to EyeWire. As a VCS game, they noted that there was no end goal at which point players could feel they had beaten the game - no final boss to defeat or challenge to overcome.

> "The ultimate reward systems, in some games, is to give the feeling of having beaten something... Almost no citizen science game is going to give that feeling."

While the scientific research aims associated with VCS serve as end goals, the team noted that research moves far slower than many players had first assumed. As a result, in early interactions between players and researchers, the players became disheartened at the slow rate with which research was progressing.

> "One of the players was like 'so what have you learned from our neurons since last time'? And I think one of the researchers was like... 'Oh, we haven't even looked at them'... It made the player feel like crap, because 'oh man, we've been working so hard!'... So in a way, to ease the burn of the slowness."

Furthermore the highly challenging and potentially abstract nature of some final research goals (for example "curing cancer") makes them unsuitable for introduction as final 'end point' challenges. Due to this, the team were unable to introduce more diegetic game elements such as narratives and characters - a factor they felt lead to further issues, such as a lack of progression.

> "One thing that's important for some games is to know you're getting better at it... [In Eyewire] there isn't a level system. You aren't playing level 20 because there is no level 20."

In essence, they were unsure how to make the game 'fun'. These difficulties were highlighted during the launch of the game, when despite initial interest in the project, contribution levels quickly fell.

> "Because we launched Eyewire, there was a huge bump in traffic and then after like, a month it just dwindled back down and we were like 'crap, what's going on?'"

The team concluded that this was due to the relatively basic nature of the game and a lack of gaming and social elements. At the time, the Eyewire interface was relatively basic. Players had access to the live chat function, cube tracing tool and a single, simplified leaderboard which reset at the end of the day, through which players could track their point scores.

Looking at the areas they felt to be strengths of citizen science games, the team made the decision to leverage the social and community aspects of the game. From this, they identified competition and associated events as a key design area that was lacking from Eyewire. Although the game had some competitive elements such as points and leaderboards, the team felt that these were insufficient for attracting participation without competition to contextualise and offer meaning to point scores.

> "If Mario were going along just collecting coins it would sort of be like 'woo, he's getting coins? Who cares! The coins don't... translate into anything!"

More specifically, given the lack of opportunities for overarching narratives and progression, these competitive elements were to form the basis of converting the Eyewire task into a game:

> "There were so many elements of Eyewire that were missing. There was no competition... You know, a game

Table 1: Competition Type.

Type	Description	Availability
Accuracy Happy Hour	Achieve accuracy above 80%	Major
Evil Cubes	Complete 12 very difficult cubes	Major
Happy Hour	Receive bonus for performance over two hours	Scheduled, Special
Marathon	Collaboratively complete cell within 24 hours	Minor, Major
Team Versus	Earn most points within time limit as team	Minor, Major
Trivia	Answer chat questions quickly and accurately	Major

[like Eyewire] without competition is arguably not a game. We had a leaderboard, but that leaderboard reset every day."

Leaderboards were expanded to allow players to track their performance over periods of days, weeks and months relative to other players. More significantly, however, players were invited to take part in ad-hoc competition challenges, where individual high-performing players would challenge the community to identify the best EyeWire tracer. Given the positive response to these contests, the team expanded the competitions, introducing regular team-based point generating challenges and regular collaborative timed cell completion challenges. Ultimately, so-called *major* competitions were introduced: week long events during which different competition challenges are held within an overarching diegetic narrative. Players must complete competition challenges to advance the narrative and overcome obstacles - for example, receiving clues to the identity of a killer in a whodunnit murder mystery themed event.

In contrast with minor competitions, major competitions are held less regularly (quarterly rather than bi-weekly/monthly) and offer specific rewards such as prizes and limited edition badges which players can display on their player profile. While minor competitions have a theme (for example, analogue vs digital) this theme expands only to player team names and a blog posting accompanied by artwork. Major competitions receive fully fleshed out narratives, with blog artwork and characters and in-game thematic elements such as in character chat communication from game masters. There is no difference in the core gameplay of minor and major competition variants, or in how long the individual competition is held for. A summary of competition types can be seen in Table 1.

Desired outcomes

Unlike some other citizen science projects, where existing work is carried out by algorithms (e.g., (Khatib et al. 2011)) or small science teams (e.g., (Lintott et al. 2008)), the research team predominantly relied on paid professionals to trace neurons prior to Eyewire. The success of EyeWire is therefore predominantly judged in terms of cost-effectivenss and efficiency - that is, is it more efficient to gather data through the large-scale crowdsourcing process rather than paying individual professionals? Initially the team found it difficult to improve on the performance of the paid professionals, noting three key issues. The first is that the tracing task is complex and takes time for individuals to learn to perform traces at a high enough level of accuracy for use in research. Two further concerns relate to the number of contributors and the efficiency with which players contribute to EyeWire. As with other citizen science platforms, a small minority of players do the majority of the work (Ponciano et al. 2014; Tinati et al. 2015a).

In the case of EyeWire, however, a minority of specially selected top-performing players must not only complete tracing tasks, but also administrative tasks to maintain site functionality platform. Dubbed 'scythes', these players are tasked with correcting inaccuracies in player traces and marking cubes as either completed or in need of further work. Without this work, the underlying algorithms may generate seeds for players, known as 'mergers', leading to wasted trace effort. Yet despite dedicating large volumes of time to the platform, even these highly active users are also fairly inefficient, dedicating a majority of their time to the IM chat interface rather than tracing (Tinati et al. 2015a).

Given these challenges, the first outcome of introducing competitions anticipated by the team was an increase in the number of EyeWire players contributing cubes to the project. In particular, the team have focussed their efforts on major competitions which feature limited edition 'badge' rewards which players can display in their in-game profile, occasional EyeWire branded merchandise prizes and the use of e-mail to notify players, particularly lapsed and inactive players - those who had been inactive for months or longer - of upcoming major competitions. From these we derive two hypotheses:

H.1 - Competitions lead to an increased number of participants compared with non-competition periods.

H.2 - Major competitions attract more players than their minor counterparts.

A second and perhaps unsurprising desired outcome is to increase overall effiency and the number of cubes submitted by players. While the team were generally optimistic about the effect of competition activities on cube contribution rates, they nevertheless noted the existence of a group of 'lurkers' - a proportion of the EyeWire player base who contribute in a steady manner and are unaffected by competitions or other efforts to increase their participation in the game. For this reason, the team stated their opinions that competition effectiveness would be largely dependent on whether players chose to participate or not.

H.3 - Competitions lead to increased cube contributions over non-competition periods.

H.4 - Levels of cube contribution from non-participating players are unaffected by competitions.

The third outcome described by the participants was to increase participation within the IM chat interface. When introducing competitive elements, the team also strengthened social aspects of the project to create an engaging game atmosphere. The most important of these elements was use of the IM chat interface. The team described two key aims of chat participation: increasing participation in EyeWire by providing an interruption from task completion which could otherwise prove monotonous and improving player performance by answering queries and through specially selected

Table 2: Daily mean statistics for Each Competition class (Dataset 1) * - Chat statistics unavailable for happy hour extra class due to a lack of instances of this class.

Type	Players	Percent Competing	Cubes	Chat Participants	Chat Messages	Percent of messages from competitors
Acc HH	238.67	25.24%	5395.83	78.50	1606.00	70.21%
Happy (Sched)	153.61	34.94%	4179.94	41.57	746.35	78.29%
Happy (Extra)	133.15	38.44%	3971.08	N/A*	N/A*	N/A*
Marathon Min	146.75	35.05%	6300.75	37	766	87.86%
Marathon Maj	164.6	40.74%	5996.63	31	506	75.60%
Team Vs Min	152.59	36.58%	4429.09	45.36	955.20	77.75%
Team Vs Maj	191.20	29.05%	5887.38	43.80	911.40	81.37%
None	139.18	N/A	3503.75	27.07	360.61	N/A

player 'mentors' with access to tools to help new and inexperienced players. Major competitions make use of use the chat interface to provide thematic hints and to hold the *trivia* competition type, which is unique to major competition periods. However, since these aspects are absent from minor competitions, the team were sceptical that chat participation would increase substantially outside of major periods.

H.5 - Only major competition periods generate increased numbers of chat messages.

H.6 - Only major competition periods lead to increased numbers of players participating in chat.

Statistical Analysis

Player Numbers In contrast to the expectations of the Eyewire team, only a small minority of EyeWire players directly participate in competition events. Within a one year period, a total of 10,296 players contributed at least 1 cube to the EyeWire project. However, only 494 players earned at least one point, or contributed at least one cube to one of the 143 competitions held during this time, a figure which represents just 4.80% of the active players during this year. A summary of player statistics for the different competition types can be seen in Table 2. With the exception of the "extra" happy hour class, the mean number of players is higher for all competition days than for days on which competitions are not held. It should be noted that although the Accuracy Happy Hour appears to be the most popular competition, we believe that this is largely due to the day on which it is held, which is always the first day of the major competitive period. This may explain the large number of active players, as far from being a highly popular competition, the accuracy happy hour attracts on average 13 players.

Competitions nevertheless have a significant effect on the number of active players in Eyewire. On days when competitions are held, player numbers are moderately higher than non-competition days (d=0.55, p$<$0.01). Major competitions are also seen to be much more effective in attracting players than minor competitions (d=0.88, p$<$0.01), in line with H.2 as suggested by the Eyewire team. In fact, our analysis shows that this is predominantly due to a proportion of players who do not participate in competitions yet choose to sign in and contribute on days when competitions are held. Not only is there a statistically significant increase in active players from the non-competing group on competition days (A1, d=0.55, p$<$0.01), but this effect is greater during major competition periods over minor periods (A2, d=0.92, p$<$0.01). On the other hand, although these increases are statistically significant, these represent a small minority of the 9,802 players who chose not to join any competitions.

Task Completion Players who sign up for competitions complete more work than players who are otherwise active on the platform. Competitions have a medium to large effect on the number of cubes contributed by competition participants as a group (d=0.79, p$<$0.01), as well as a small to medium effect on the number of cubes per competing player (d=0.49, p$<$0.01) as can be seen in figure 2. Nevertheless, further analysis reveals that this increase in cube contributions extends to non-participating players, who as a group display a small increase in the number of cubes submitted on days where competitions are held (A3, d=0.35, p$<$0.01), particularly on major competition days (A4, d=0.83, p$<$0.01). In terms of competition types, cube contributions vary greatly. While all competition types encourage greater numbers of cube submissions, the Marathon class attracts higher numbers of cubes in comparison with other classes, particularly the minor marathon class.

Community Discussion An analysis of the 91 day sample of chat participation reveals a notable increase in chat participation during competitions. When comparing days on which competitions occur with those on which no competitions are held, test statistics identify a moderate increase in the total number of chat participants (d=0.60, p$<$0.01). Although the majority of the mechanisms introduced by the Eyewire team to motivate chat participation were focused on major competitive periods, both major and minor competitions demonstrate significant increases in the numbers of chat messages and numbers of messages sent per player (p$<$0.01). Nevertheless, this effect is larger for major competitions (d=0.85) than for minor competitions (d=0.62). However, when comparing the number of chat participants on minor and major competition days with non-competition days, the analysis was inconclusive.

Discussion

Our results show that while competitions appeal to a minority of human computation contributors, they nevertheless have a significant positive impact on the number of contributions generated, as well as the level of community discussion activity. This effect is in part due to the effect that competitions have on those participants who choose not to actively compete, yet who nonetheless choose to sign-in and contribute more during competition periods. Our findings suggest that competitions have an overall positive effect on participation in both chat and task activities. We note, however, that EyeWire is a highly gamified context and suggest that further analysis is required to understand the use of competitions in non-gamified human computation initiatives - particularly those which rely solely on altruistic and intrinsic motivations.

Table 3: Summary of Hypothesis, *H*, and additional Mann-Whitney U test, *A*, outcomes. Z-crit for p<0.01 = -2.33

H	Condition	Z-Stat	*d*	Conclusion
H1	Number of players (competition vs non-competition)	-5.1	0.55	Hypothesis confirmed (p<0.01, medium effect size)
H2	Number of players (minor vs major)	-3.5	0.88	Hypothesis confirmed (p<0.01, large effect size)
H3	Total cube contributions (competition vs non-competition) Cube contributions per player (competitions vs non-competition)	-6.99 -4.58	0.79 0.49	Hypothesis confirmed (p<0.01, medium/large effect size) Hypothesis confirmed (p<0.01, small/medium effect size)
H4	Total cube contributions for non-competing players (competitions vs non-competition)	-3.33	0.35	Hypothesis incorrect. Cube levels from non-competing players are higher during competitions (p<0.01, small effect size)
H5	Chat messages per player (major vs non-competition) Chat messages per player (minor vs non-competition)	-3.08 -2.37	0.85 0.62	Hypothesis incorrect. Both minor and major competitions generate increased chat messages compared with non-competition days (p<0.01, large, medium effect size)
H6	Total chat participants (competitions vs non-competition)	-2.64	0.60	Uncertain. Results show an increase in chat participants over all competition types (p¡0.01, medium effect size)
A1	Numbers of non-competing players (competition vs non-competition)	-5.1	0.55	Competitions result in a significant increase in non-competing player numbers (p<0.01, medium effect size)
A2	Numbers of non-participating players (minor competitions vs major competitions)	-3.65	0.92	Major competitions attract significantly more non-competing players than minor competitions (p<0.01, large effect size)
A3	Total cube contributions from non-competing players (competitions vs non-competitions)	-3.33	0.35	During competitions, cube contributions from non-competing players increase (p<0.01, small effect size)
A4	Total cube contributions from non-competing players (minor competitions vs major competitions)	-3.36	0.83	During major contributions, more cubes are contributed by non-competing players than during minor competitions (p<0.01, large effect size)

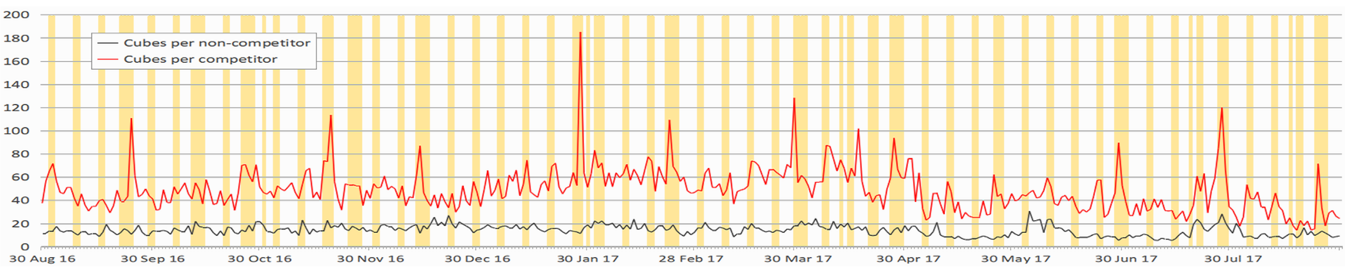

Figure 2: Cubes per non-competing and competing players on competition (highlighted) and non-competition days.

Major competitions attract significantly more competitors, non-competing players and cube contributions than minor competition events. Although the team predicted an increase in competitor and cube numbers, it is nevertheless surprising that major competitions should be significantly more effective than minor competitions at encouraging cube contributions. While major events feature limited edition badge rewards for participation, these are given to all players who reach the (relatively low) threshold of points or cubes - dependent on competition type - and there are no other mechanisms to encourage or reward contributions that cannot also be earned in minor competitions. Previous studies have demonstrated that participants in human computation initiatives are more motivated and engaged by diegetic, narrative and progression-driven gamified experiences, even when the underlying task itself is identical (Prestopnik and Tang 2015). Our findings suggest that this may be true even in circumstances where the elements are more implicit and community-driven - that is, in the case of EyeWire, where the only visible elements of the progression and narrative are visible through the chat or blog features, such as 'whodunnit' clues.

Similarly, although our findings demonstrate an increase in non-competitor numbers and participation levels during competitions, the reason for this is unclear. EyeWire features few opportunities for spectation or collaboration with other players and competition rewards are provided only to those who participate. One possibility, as raised by the team, is the potential impact of a positive chat atmosphere - as the team stated: *"chat is blowing up. Everyone's excited."* Moreover, it should be noted that the chat and community-driven narrative features are still visible to these non-competing players, but only during major competitions - a factor that would explain the significant increase in non-competitor numbers and contributions during the major competition class. We believe this is an interesting area for further study and that the inclusion of more visible competition elements that include the whole community may be important for maximising the effectiveness of competitions in human computation initiatives. A further possibility is that the increase in non-competitor numbers is a result of notification campaigns (particularly e-mails) attached to competitions - particularly since this effect is greater during major competitions, where such campaigns are held. Our data was insufficient to test such a hypothesis, as we lack confirmation that players had received or seen campaign emails (typically sent shortly before competitions occur) and whether notifications or competitions - or both - drive any return to EyeWire. Nevertheless, our findings show that any impact on participation is extremely brief, coinciding only with competitions (see figure 2). Moreover, this does not explain the increase in participation levels from within this group during competitions.

Within the literature, autonomy has been highlighted as a key motivational affordance across a variety of human computation initiatives (Lukyanenko, Parsons, and Wiersma 2016; Prestopnik and Tang 2015). However, our findings note a potential trade-off with regard to autonomy in competitions. The team versus competition class - during which participants can earn points from any cube and, where applicable, bonuses, tutorials and promotion-dependent activities - attracted a greater number of competitors, but on average resulted in reduced numbers of cube contributions when compared with the marathon class. While the marathon class is significantly less autonomous, accepting only cube contributions and only for one cell, this class has a collaborative, goal-oriented aspect which is not present in the team versus competition. Rather than earning points relative to the performance of other players, players are instead scored based on their own number of cubes contributed and must as a community reach the required number of cells within the time-limit. We note that the effectiveness of competitive and collaborative elements remains a controversial topic within the wider literature and we suggest that further exploration is required to understand how autonomy fits within this space.

Further related to autonomy, a number of VCS studies have made explicit or implicit connections between autonomy and the chat platforms, suggesting that chat participation can replace or overcome a lack of autonomy in tasks (Jackson et al. 2014; Reeves et al. 2017). This view was even put forward by the EyeWire team: *"chat makes up for the dryness of the task."* On the other hand, our results suggest that this may not be the case, at least in EyeWire. During the less autonomous marathon class, fewer chat messages were sent by the community and the number of chat participants was lower than during the much more autonomous team versus category. This suggests that while more intensive, time-limited challenge activities may be effective for motivating task completion, these restrictions discourage autonomous activity and engagement with other tasks such as chat.

Finally, a key area for research, which we have not addressed in our analysis, is consideration of the impact of competitions on the quality and accuracy of responses. The nature of EyeWire makes it unsuitable for such analysis, as the platform largely lacks ground truth cubes, assesses responses through majority voting and includes an iterative process through which cubes can be corrected and marked for further analysis by the promoted 'scythe' player class (Reeves et al. 2017). In contrast with the view held by the team that competitions can effectively train players to contribute to human computation tasks, our findings demonstrate that the accuracy happy hour class attracts only a small number of participants. However, this does not discount the possibility of players learning through competitions, as it takes many hours and thousands of cubes to 'master' EyeWire and as such, the increase in activity that competitions encourage is crucial for newer and less active players.

In addition, further consideration must be given to additional methods for overcoming the weaknesses and challenges of VCS and associated game genres. This may include testing other progression formats - for example, the EyeWire team suggested opportunities posed by badges or unlockables, offering individual-centred rather than communal progression. Opportunities for team or group play are also of interest, given the apparent importance of sociality in VCS, but we caution that such opportunities also pose issues for common quality assurance processes such as redundancy (Wiggins et al. 2011). Given the uncommon and temporary nature of competitions across VCS (Reeves et al. 2017), it is unclear to what extent our findings generalise, particularly to less socially-driven or gamified projects where participant motivations differ (Mekler et al. 2013b).

Limitations

Due to the nature of EyeWire's chat system and volume of messages produced, these messages are not readily archived and so we were only able to collect chat messages live from the system. For this reason, our second dataset covered only 1/4 of the larger first dataset. We noted nonetheless that this period covered 48 competition days, including a range of minor and major events and therefore felt it was suitable for analysis. To further overcome this issue, we analysed hypotheses covering minor and major competitions, rather than specific competition types, to maximise the volume of available data. We do, however, make the assumption that non-, minor and major competition activity during this 91 day period is indicative of activity in other periods.

Conclusion

Overall, this study contributes to the understanding of how competitions influence player participation in human computation activities. We have identified the underlying rationale for introducing competitions to Eyewire. Our findings demonstrate that competitions effectively increase levels of contribution for task and chat components of the project, while increasing the number of active players. This effect persisted even with those players who choose not to contribute to competitions, who exhibit greater numbers of active players and cube contributions on competition days. Based on these findings, we identified points of contention with previous human computation literature, as well as a number of areas for further research and analysis.

Acknowledgements

This research was funded by the Digital Economy programme, as part of UK Research and Innovation under grant agreement EP/G036926/1 and the Horizon2020 programme, through the Stars4All project under grant agreement 688135. We would like to thank the EyeWire team for their time and support and the anonymous reviewers for their comments and recommendations.

References

Aoki, P.; Woodruff, A.; Yellapragada, B.; and Willett, W. 2017. Environmental protection and agency: Motivations, capacity, and goals in participatory sensing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3138–3150. ACM.

Borrett, S., and Hughes, L. 2016. Reporting methods for processing and analysis of data from serial block face scanning electron microscopy. *Journal of microscopy* 263(1):3–9.

Bowser, A.; Hansen, D.; He, Y.; Boston, C.; Reid, M.; Gunnell, L.; and Preece, J. 2013. Using gamification to inspire new citizen science volunteers. In *Proceedings of the first international conference on gameful design, research, and applications*, 18–25. ACM.

Brouwer, R. 2016. When competition is the loser: the indirect effect of intra-team competition on team performance through task complexity, team conflict and psychological safety. In *System Sciences (HICSS), 2016 49th Hawaii International Conference on*, 1348–1357. IEEE.

Cheng, J.; Teevan, J.; Iqbal, S. T.; and Bernstein, M. S. 2015. Break it down: A comparison of macro-and microtasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 4061–4064. ACM.

Cox, A.; Cairns, P.; Shah, P.; and Carroll, M. 2012. Not doing but thinking: the role of challenge in the gaming experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 79–88. ACM.

Cox, J.; Oh, E. Y.; Simmons, B.; Graham, G.; Greenhill, A.; Lintott, C.; Masters, K.; et al. 2015. Doing good online: An investigation into the characteristics and motivations of digital volunteers.

Curtis, V. 2014. Online citizen science games: opportunities for the biological sciences. *Applied & translational genomics* 3(4):90–94.

Deterding, S.; Dixon, D.; Khaled, R.; and Nacke, L. 2011. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, 9–15. ACM.

Dickinson, J. L.; Zuckerberg, B.; and Bonter, D. N. 2010. Citizen science as an ecological research tool: challenges and benefits. *Annual review of ecology, evolution, and systematics* 41:149–172.

Eveleigh, A.; Jennett, C.; Lynn, S.; and Cox, A. L. 2013. "i want to be a captain! i want to be a captain!": Gamification in the old weather citizen science project. In *Proceedings of the first international conference on gameful design, research, and applications*, 79–82. ACM.

Greenhill, A.; Holmes, K.; Woodcock, J.; Lintott, C.; Simmons, B. D.; Graham, G.; Cox, J.; Oh, E. Y.; Masters, K.; and Lewandowski, D. 2016. Playing with science: exploring how game activity motivates users participation on an online citizen science platform. *Aslib Journal of Information Management* 68(3).

Haklay, M. 2013. Citizen science and volunteered geographic information: Overview and typology of participation. In *Crowdsourcing geographic knowledge*. Springer. 105–122.

Iacovides, I.; Jennett, C.; Cornish-Trestrail, C.; and Cox, A. L. 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*, 1101–1106. ACM.

Jackson, C.; Østerlund, C.; Crowston, K.; Mugar, G.; and Hassman, K. 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.

Jennings, N. R.; Moreau, L.; Nicholson, D.; Ramchurn, S.; Roberts, S.; Rodden, T.; and Rogers, A. 2014. Human-agent collectives. *Communications of the ACM* 57(12):80–88.

Jia, Y.; Xu, B.; Karanam, Y.; and Voida, S. 2016. Personality-targeted gamification: a survey study on personality traits and motivational affordances. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2001–2013. ACM.

Jia, Y.; Liu, Y.; Yu, X.; and Voida, S. 2017. Designing leaderboards for gamification: Perceived differences based on user ranking, application domain, and personality traits. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1949–1960. ACM.

Khatib, F.; Cooper, S.; Tyka, M. D.; Xu, K.; Makedon, I.; Popović, Z.; Baker, D.; and Players, F. 2011. Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences* 108(47):18949–18953.

Kim, J. S.; Greene, M. J.; Zlateski, A.; Lee, K.; Richardson, M.; Turaga, S. C.; Purcaro, M.; Balkam, M.; Robinson, A.; Behabadi, B. F.; et al. 2014. Space-time wiring specificity supports direction selectivity in the retina. *Nature* 509(7500):331.

Kumar, J. 2013. Gamification at work: Designing engaging business software. In *International Conference of Design, User Experience, and Usability*, 528–537. Springer.

Lieberoth, A.; Pedersen, M. K.; Marin, A. C.; Planke, T.; and Sherson, J. F. 2014. Getting humans to do quantum optimization-user acquisition, engagement and early results from the citizen cyberscience game quantum moves. *Human Computation* 1(2).

Lieberoth, A. 2015. Shallow gamification: Testing psychological effects of framing an activity as a game. *Games and Culture* 10(3):229–248.

Lintott, C. J.; Schawinski, K.; Slosar, A.; Land, K.; Bamford, S.; Thomas, D.; Raddick, M. J.; Nichol, R. C.; Szalay, A.; Andreescu, D.; 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):1179–1189.

Luczak-Roesch, M.; Tinati, R.; Simperl, E.; Van Kleek, M.; Shadbolt, N.; and Simpson, R. 2014. Why won't aliens talk to us? content and community dynamics in online citizen science.

Lukyanenko, R.; Parsons, J.; and Wiersma, Y. F. 2016. The impact of task design on accuracy, completeness and discovery in surveillance-based crowdsourcing. In *Proceedings of the 4th International Conference on Collective Intelligence*.

Luu, S., and Narayan, A. 2017. Games at work: Examining a model of team effectiveness in an interdependent gaming task. *Computers in Human Behavior*.

Mattos, A. B.; Herrmann, R.; Shigeno, K. K.; and Feris, R. 2014. A mission-oriented citizen science platform for efficient flower classification based on combination of feature descriptors. In *EMR@ ICMR*, 45–52.

Mekler, E. D.; Brühlmann, F.; Opwis, K.; and Tuch, A. N. 2013a. Disassembling gamification: the effects of points and meaning on user motivation and performance. In *CHI'13 extended abstracts on human factors in computing systems*, 1137–1142. ACM.

Mekler, E. D.; Brühlmann, F.; Opwis, K.; and Tuch, A. N. 2013b. Do points, levels and leaderboards harm intrinsic motivation?: an empirical analysis of common gamification elements. In *Proceedings of the First International Conference on gameful design, research, and applications*, 66–73. ACM.

Michelucci, P. 2013. *Handbook of Human Computation*. Springer Science & Business Media.

Oliveira, N.; Jun, E.; and Reinecke, K. 2017. Citizen science opportunities in volunteer-based online experiments. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 6800–6812. ACM.

Østerlund, C.; Mugar, G.; Jackson, C. B.; Hassman, K.; and Crowston, K. 2014. Socializing the crowd: Learning to talk in citizen science. In *Academy of Management Annual Meeting, OCIS Division, Philadelphia, PA*.

Ponciano, L.; Brasileiro, F.; Simpson, R.; and Smith, A. 2014. Volunteers' engagement in human computation for astronomy projects. *Computing in Science & Engineering* 16(6):52–59.

Ponti, M.; Hillman, T.; and Stankovic, I. 2015. Science and gamification: The odd couple? In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 679–684. ACM.

Prestopnik, N. R., and Tang, J. 2015. Points, stories, worlds, and diegesis: Comparing player experiences in two citizen science games. *Computers in Human Behavior* 52:492–506.

Prestopnik, N.; Crowston, K.; and Wang, J. 2014. Exploring data quality in games with a purpose. *iConference 2014 Proceedings*.

Reeves, N.; Tinati, R.; Zerr, S.; Van Kleek, M. G.; and Simperl, E. 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*, 2137–2152. ACM.

Rotman, D.; Preece, J.; Hammock, J.; Procita, K.; Hansen, D.; Parr, C.; Lewis, D.; and Jacobs, D. 2012. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, 217–226. ACM.

Sandbrook, C.; Adams, W. M.; and Monteferri, B. 2015. Digital games and biodiversity conservation. *Conservation Letters* 8(2):118–124.

Sauermann, H., and Franzoni, C. 2015. Crowd science user contribution patterns and their implications. *Proceedings of the National Academy of Sciences* 112(3):679–684.

Sawilowsky, S. S. 2009. New effect size rules of thumb. *Journal of Modern Applied Statistical Methods* 8(2):26.

Seaborn, K., and Fels, D. I. 2015. Gamification in theory and action: A survey. *International Journal of Human-Computer Studies* 74:14–31.

Siu, K.; Zook, A.; and Riedl, M. O. 2014. Collaboration versus competition: Design and evaluation of mechanics for games with a purpose. In *FDG*.

Sprinks, J.; Houghton, R.; Bamford, S.; and Morley, J. 2015a. Citizen scientists: The importance of being needed and not wasted. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play. ACM*.

Sprinks, J.; Houghton, R.; Bamford, S.; and Morley, J. 2015b. The impact of task workflow design on citizen science users and results. *Contemporary Ergonomics and Human Factors* 371–378.

Tinati, R.; Luczak-Rösch, M.; Simperl, E.; Shadbolt, N.; and Hall, W. 2015a. '/command'and conquer: Analysing discussion in a citizen science game. In *Proceedings of the ACM Web Science Conference*, 26. ACM.

Tinati, R.; Van Kleek, M.; Simperl, E.; Luczak-Rösch, M.; Simpson, R.; and Shadbolt, N. 2015b. 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*, 4069–4078. ACM.

Tinati, R.; Luczak-Roesch, M.; Simperl, E.; and Hall, W. 2017. An investigation of player motivations in eyewire, a gamified citizen science project. *Computers in Human Behavior* 73:527–540.

Tinati, R.; Simperl, E.; and Luczak-Roesch, M. 2017. To help or hinder: Real-time chat in citizen science. *Proceedings of ICWSM* 2017.

Wiggins, A., and Crowston, K. 2011. From conservation to crowdsourcing: A typology of citizen science. In *System Sciences (HICSS), 2011 44th Hawaii international conference on*, 1–10. IEEE.

Wiggins, A.; Newman, G.; Stevenson, R. D.; and Crowston, K. 2011. Mechanisms for data quality and validation in citizen science. In *e-Science Workshops (eScienceW), 2011 IEEE Seventh International Conference on*, 14–19. IEEE.

Zheng, H.; Li, D.; and Hou, W. 2011. Task design, motivation, and participation in crowdsourcing contests. *International Journal of Electronic Commerce* 15(4):57–88.