Supporting Information

Sauermann and Franzoni 10.1073/pnas.1408907112

SI Text

Project Characteristics and Key Measures. Table S1 summarizes key characteristics of the seven projects including, among others, the high-level objective of the analysis, the type of raw data provided for inspection (e.g., image, video), the activity that participants are asked to perform, the more fundamental cognitive task involved in performing these activities (1), and some of the common disturbances that make the cognitive task ambiguous (and thus require human intelligence). We also note the start date of each project and the end of our observation period (180 d after the start date). Note that, although six projects operated continuously, the project Galaxy Zoo Supernovae had some days on which it ran out of data and stopped accepting contributions (see also Fig. S4, which shows 15 d with zero activity). We decided to keep this project in the analysis because it is a real project and provides interesting data points. However, statistics concerning this particular project should be interpreted with this particularity in mind. Another particularity is that Old Weather uses some aspects of gamification in that users can earn different ranks (e.g., Lieutenant, Captain) based on their number of classifications. A prior interview-based study of Old Weather suggests that some users like this feature, whereas others dislike it, with no clear overall tendency (2). Because only one of the projects uses gamification, we cannot empirically test the effects of this feature.

Classifications per day. The output of processing one object in Zooniverse projects is called a classification. Table S1 indicates the particular activities performed for a classification in each project. The data used in this study include a count of classifications completed by each person for each day.

Time spent per day. The time spent by a contributor on a given day was computed by Zooniverse as the difference between the time of the last classification and the time of the first classification recorded on that day. Because participants may have stopped working between two classifications, the clock stops after 30 min without a classification; classifications before this break and classifications made after this break are considered parts of two separate sessions within a given day. In that case, the total time per day is computed as the sum of the durations of the separate sessions. A limitation of this time measure is that the time recorded for user-days with only one classification is zero (~12% of user-days). To mitigate this problem, we compute the average time per classification for each contributor based on data from contributor-days with multiple classifications and use the median of this value (across all users in a project) as the best estimate of the time users spent on user-days with only a single classification. This adjustment changes estimates of total time contributed by less than 1% (from 128,487 to 129,540 h).

Analyses. In the following, we provide details on the analyses reported in the main text (in the order in which they appear there), as well as a number of supplementary analyses.

Estimate of counterfactual cost of labor using hourly wages. We multiply the number of hours of effort received by each project with the typical hourly wage of an undergraduate research assistant in the United States. Because no standard wage exists, we estimated this wage as roughly $12 based on information aggregated at www.glassdoor.com/Salaries/undergraduate-research-assistant-salary-SRCH_KO0,32.htm, as well as information available on the websites of US universities (e.g., www.utexas.edu/hr/student/compensation.html; www.washington.edu/admin/hr/ocpsp/student/; and www.ohr.wisc.edu/polproced/utg/SalRng.html). This information was accessed on July 25, 2014. Undergraduate hourly wage rates are a lower bound for the cost of labor in an academic research laboratory, and the costs of graduate students and postdocs are likely to be significantly higher (3). Because the tasks performed by volunteers in Zooniverse crowd science projects do not require PhD level training, however, undergraduate wages provide the most reasonable (and relatively conservative) counterfactual cost estimate.

For readers wishing to apply different rates, including annual costs of certain types of positions, Table S2 also provides an estimate of the number of full time equivalents (FTEs) that would be required to supply the same number of hours over 180 d. To compute this number, we assume 8 h per work day and 5 work days per week and compute the FTE for a given project as FTE = total hours worked/[8 × 180 × (5/7)]. Using this measure, volunteers did the work of more than 125 FTE. Of course, although we can convert the total number of hours contributed by volunteers into FTE, it is not clear whether 125 workers could be found that are willing to code images for 8 h each work day. Moreover, given the rather monotonous and relatively simple tasks, such full-time workers might experience exhaustion and low job satisfaction (4–6). Thus, distributing a large volume of work among many people may not only reduce the time required to complete the overall project but may also avoid fatigue or exhaustion, and make the job more fun for everyone. At the same time, repetition may lead to learning and increased efficiency (see below). These and other potential tradeoffs from using crowd labor vs. traditional full-time employees seem a particularly fruitful area for future theoretical and empirical work.

Estimate of counterfactual cost of classifications using AMT pricing. We multiplied the number of all classifications contributed to a project with the estimated market price of one classification. The latter was determined based on pricing information collected from Amazon Mechanical Turk (AMT) (https://www.mturk.com/). AMT is an online crowdsourcing platform that is currently considered the largest intermediary for tasks requiring human intelligence and has also been used extensively for research on crowdsourcing (7, 8). We browsed the catalog and examples of tasks and used the prices suggested by the platform for the closest possible task. Price information was accessed and retrieved on February 13, 2014. The price suggested for complex image tagging and for image transcription on AMT is $0.05. There is no single suggested price for image categorization in AMT, presumably because the effort required could vary considerably depending on the complexity of the image and the number of categories provided. However, AMT discourages setting prices below $0.02 per categorization. Given that the examples of categorization provided on AMT are simpler than those typical of Zooniverse projects, but less time-consuming than the typical AMT image transcription, we set the unit price for categorizations to an intermediate value of $0.035. The kind of video categorization that Solar Stormwatch required in 2010–2011 (participants were asked to watch a video, tag the start and end point of a solar explosion using a still-video tool, and provide classifications) has no immediate equivalent in AMT. We therefore chose to apply pricing suggested on AMT for a short video transcription ($1). The following list summarizes the assumptions made to estimate current market prices for one classification in each project. The resulting counterfactual costs per classification and for the total contributions made to each project are listed in Table S2.

- Solar Stormwatch: Watch video of ∼1 min, classify and tag; at $1 each
- Galaxy Zoo Supernovae: Approximately three categorizations per image at $0.035 each
- Galaxy Zoo Hubble: Approximately four categorizations per object at $0.035 each
- Moon Zoo: Approximately five simple tags per image at $0.035 each
- Old Weather: Approximately 13 transcriptions per object (1 trans. of date, 1 trans. of location; 1 trans. of fuel consumption; ∼2 trans. of wind direction; ∼4 observations of temperature; ∼4 observations of pressure) at $0.05 each
- Milkyway Project: Approximately three tags per image at $0.05 each
- Planet Hunters: Approximately three categorizations at $0.035 each and one tag per image at $0.05

Although AMT provides useful counterfactual cost estimates for procuring classifications via online labor markets, we cannot tell how projects of the scale studied here would perform on AMT. Indeed, given the differences in infrastructure, incentive systems, and possibly composition of the crowd (9), contribution dynamics and project performance may be quite different. As such, future research studying whether and how the same scientific problem can be solved using different crowd-based mechanisms and platforms would be particularly interesting.

Lorenz curves and Gini coefficients (Fig. 3). The Lorenz curves shown in Fig. 3 plot the cumulative share of total classifications (y axis) made by a particular cumulative share of users (x axis). The 45° line indicates total equality, i.e., all users contribute equally. The stronger the curvature of the Lorenz curves, the stronger the inequality in contributions. We computed the Lorenz curves using the statistical software package Stata. Although all projects have Lorenz curves that are clearly different from the 45° line, they also differ significantly from each other. We tested the equality of the distributions using Kolmogorov-Smirnoff tests and found that all distributions are significantly different from each other at the 5% level of confidence.

The Gini coefficient reflects the ratio of the area between the 45° line and the Lorenz curve for a particular project on the one hand and the total area under the 45° line on the other. If all contributions are equal, the Lorenz curve and the 45° line overlap perfectly, leading to a Gini coefficient of 0. If one contributor makes all of the contributions, the Gini coefficient is 1. As such, higher Gini coefficients indicate higher concentration in contributions.

Average time per classification. Average time per classification was computed by dividing the total time spent on a project by the total number of classifications made by a particular individual.

Speed advantage of top contributors vs. non-top contributors. To explore whether users in the top 10% in terms of total classifications work faster than others, we compared the average time per classification for top contributors to that of non–top contributors. For each project, Table S2 shows the difference between the two numbers expressed as a percentage of the time per classification for non–top contributors.

Changes in speed over time. Our finding that top contributors (top 10% in terms of classifications in a particular project) work somewhat faster than non–top contributors raises the question of whether higher speed reflects some innate ability advantage of a person (i.e., it is fixed) or whether it emerges over time (e.g., due to learning). To explore potential changes over time, we focus on top contributors that have at least 7 active days in a project (across projects, 82.75% of individuals with at least 7 active days are also top contributors, and 41.01% of top contributors have at least 7 active days). This sample includes a total of 4,083 individuals, ranging from 130 users in the smallest project (Galaxy Zoo Supernovae) to 1,552 in Planet Hunters. For each of these users, we compute the average time per classification for each of the first 7 active days and average across users to obtain the average speed for a given day at the project level. To make measures comparable across projects, we then index the time per classification to 100% for the first day and express time per classification on subsequent days relative to that of the first day. Fig. S1 plots the results. We observe that speed increases over time in all of the projects, with the reduction in time per classification ranging from roughly 20% to 37%, consistent with learning effects. Moreover, the increase in speed seems most pronounced early on (between days 1 and 3) and then continues at a smaller rate. To formally test these changes, we estimate a series of regression models. In particular, we use the same subsample of individuals and estimate OLS models that regress the time per classification for each of the first 7 active days on a dummy variable indicating the day number. Because we use seven observations per individual, we can include individual fixed effects to control for unobserved heterogeneity. As such, these regressions show how classification speed changes as a given individual progresses from active day 1 to active day 7. The results confirm a significant increase in speed, as reflected in significant negative coefficients of the day dummies (Table S3).

Although these analyses show that the classification speed of top contributors increased over their tenure in the project, they do not rule out the possibility that top contributors were also faster in the beginning. To examine this possibility, we compute the average time per classification for all individuals on their first day and compare speed between top and non–top contributors. We make the interesting observation that top contributors do not exhibit a clear speed advantage on their first day. Rather, in four of seven projects, they tend to be slower on their first day than non–top contributors (resulting in an average of −11% across the seven projects; Table S2). One potential interpretation is that individuals with a stronger interest in a project (i.e., latent top contributors) are more willing to invest in learning by working slowly and taking the task more seriously than others. It is interesting that the only project where top contributors have a sizeable advantage on the first day is Galaxy Zoo Hubble, which is very similar in nature to the original Galaxy Zoo projects (started before 2010). Although we do not have data on these earlier projects, it is conceivable that some of the top contributors to Galaxy Zoo Hubble were previously active in other Galaxy Zoo projects and that some of their learning carried over to Galaxy Zoo Hubble.

Although the data do not include measures of the quality or accuracy of classifications, our observations regarding speed increases over time suggest the relationships between speed, accuracy, learning, and top contributor status as an interesting area for future research.

Share of contributors who return. Users may participate in a project only for 1 d or may return for additional days. Fig. S2 shows the distribution of active days for each project, including the share of users who participates only once (nonreturning users).

Duration of breaks between active days. For users with at least 7 active days, we computed the time between active days 1 and 2, as well as between active days 6 and 7. Table S2 shows that this time increases from an average of 5.23–8.30 d, suggesting a declining frequency of activity even for these highly active contributors. Note that the increase is even larger (from 4.05 to 8.14 d) when we exclude Galazy Zoo Supernovae, which is an outlier likely due to the fact that it did not accept contributions on all days. Due to space limitations, Table S2 does not list the breaks between other pairs of active days. Across projects, these breaks average 4.47 d between active days 2 and 3, 4.68 d between days 3 and 4, 5.11 d between days 4 and 5, and 5.51 d between days 5 and 6. Excluding Galaxy Zoo Supernovae, the breaks average 3.83 d between active days 2 and 3, 3.94 d between days 3 and 4, 4.72 d between days 4 and 5, and 5.58 d between days 5 and 6.

Average duration of daily effort. For this analysis, we take all active contributor-days for a given contributor (conditional on at least one classification) and compute the average time spent per day. Fig. S3 shows the distribution of the average duration of daily effort for each project, and Table S2 reports the means.

Defining four groups of users (Fig. 4). We classify contributors along two dimensions. The first dimension reflects whether a contributor participates only for 1 d vs. for multiple days. The second dimension reflects the user's average duration of daily effort (see above). For the second dimension, we chose as cutoff the 90th percentile across all contributors to the project. We chose this particular cutoff (rather than the median or mean time) to more clearly examine the contributions of highly active contributors.

Characterizing returning users. Given the large share of contributions made by those users who contributed to a project multiple times, we performed additional analyses to characterize return users in more detail. Given the data available, we focus on the timing of users' joining the project, using three different independent variables (Table S4). The variable start day captures on which day of the project's life a user joined (starting with 1 for users who joined on the day the project came online). As an alternative, we also code a dummy variable indicating whether a user joined in the first 7 d of the project (original user, see also discussion of Fig. 5). Finally, we code whether a user joined on the day of a spike in that project's activity (see below for details on the identification of spikes; this variable is not defined in projects Galaxy Zoo Supernovae and Galaxy Zoo Hubble). We then estimate linear probability models (LPMs) using these measures to predict whether a particular user is a return user or not (logit estimation yields the same results but LPMs are easier to interpret because coefficients can directly be interpreted as change in probabilities). An obvious problem with using the full sample is right censoring, i.e., we may underestimate the likelihood of a return for users who joined later during the 180-d observation window and are thus observed for a shorter time than users who joined earlier. To address this concern, we redefine return users as those who return within 30 d of joining a project (return30) and reestimate regressions using the sample of those users who are observed for at least 30 d. Models 1–2 in Table S4 combine cases from all projects but show no significant relationship between the timing of joining and return behavior. Models 3–16 show separate regressions for each project. We observe positive coefficients of start day in three of the projects, suggesting that users who joined later are more likely to return during the subsequent 30 d. To illustrate the effect size, consider the coefficient of 0.0004 in Moon Zoo, which implies that a user who joins on day 150 is 6 percentage points more likely to return than a user who joined on the first day (compared with a baseline of 18 percentage points; Table S2). On the other hand, the coefficient of start day is negative for three other projects, indicating that later users are less likely to return. The results for original user, which dichotomizes time of joining, are similarly mixed: this variable is not significant in four of the projects, has a positive coefficient in two, and a negative coefficient in one. We also find no systematic patterns for started during spike: this variable is not significant in one project, negative in three, and positive in one. Taken together, looking across all projects, we find no systematic relationship between the timing of a users' joining a project and subsequent return behavior. We do find significant relationships when looking at projects separately, but these relationships vary both in sign and magnitude, providing no clear picture. Although we lack the data to further investigate the observed relationships, the timing of users' joining a project and their return behavior suggest a potentially fruitful avenue for future research. In particular, it may be useful to consider a model where there is heterogeneity in the general population with respect to individuals' interest in science generally, particular fields of science, or even participation in crowd science projects (10, 11). Given such heterogeneity, the mix of users attracted to a particular project may change over time and may also be affected by media attention or promotion activities undertaken by projects (similar models have been used to study the diffusion of innovations, e.g., refs. 12 and 13).

Project level dynamics (Fig. 5). Our goal in this analysis is to examine the total number of hours received by a project for each of the first 180 d of its life and to distinguish the contributions made by different groups of users. First, we define a cohort of original contributors as those contributors who joined a project in the first 7 d of its life. Second, we define a rolling window of 6 d before and including the focal day and classify users who joined during that window as new users. The remaining users fall into a third residual category. Note that during the first 7 d of a project's life, all contributors are new and original. Although being an original contributor is a fixed attribute of a person, new contributor status is kept for only 7 d. The sum of hours contributed by the three groups of users on a given day equals the total number of hours received by the project on that day.

Fig. 5 summarizes overall patterns by showing the average levels of contributions of each group of users across the seven projects. Note that the x axis (age of project) ranges from 1 to 180 d; this analysis time is not the actual calendar time because age = 1 for each project corresponds to a different calendar date (the start date listed in Table S1). Fig. S4 shows separate graphs for each project. These graphs show that the patterns described in the main text are quite general: Total contributions are highest early in a project's life and tend to decline over time. Activity is highly variable with noticeable spikes. Contributions of original users decline rapidly and contributions from new users constitute a significant portion of total effort received by projects later in their life. The one exception is Galaxy Zoo Supernovae, which did not accept contributions continuously over the observation period.

Fig. S5 shows cumulative hours contributed to projects, distinguishing contributions by original users (i.e., those who joined in the first 7 d of project life) and contributions made by users who joined later (the distinction between new users and the third residual category is not useful for considerations of cumulative hours). To complement this figure, Table S2 lists for each project the share of total contributions made by original users, showing that across the seven projects, the users who joined in the first 7 d are responsible for roughly 33% of total hours contributed.

Complementing our analysis of the dynamics of effort contributions over time, Fig. S6 provides some insight into the dynamics of the number of users joining the project on a particular day (first time users). Note that this definition of first time users is more restrictive than the earlier analysis of new users (which joined in a rolling window of 7 d). In addition, Fig. S6 distinguishes which of these first time users are returning users (at a later point in time) vs. users who do not return. We make a number of interesting observations. First, we find that the number of first time users tends to be higher early in a project's life, consistent with the front-loaded nature of effort contributions. Second, inflows of new users are very variable over time, as reflected in sharp spikes similar to those observed in Fig. S4. Interestingly, the spikes do not correlate perfectly because some of the spikes in effort contributions were driven to a large extent by existing users rather than by an inflow of first-time users. Finally, when distinguishing between users who return during the observation window and those that do not, it appears that the share of return users is somewhat higher early in the project's life. As noted in our analysis of return users above, however, this may reflect censoring in that we observe early contributors for a longer period than those who joined toward the end of the observation window.

Spikes in activity. Fig. 5 and Fig. S4 show that contributions to projects are very volatile over time, with noticeable spikes. For exploratory purposes, we defined a "spike" as occurring when

a project receives a number of daily hours that exceed 200% of the average received in the prior 2 d. Using this definition of a spike, we observe between zero spikes (Galaxy Zoo Hubble) and seven spikes (Solar Stormwatch). Galaxy Zoo Supernovae is an outlier with 28 spikes; as discussed earlier, this project had days with zero activity, inflating the number of subsequent spikes.

Based on our discussions with project organizers, spikes can have several different causes. The high level of activity in the first few days of a project is likely attributable to the fact that Zooniverse announces new projects to a large base of existing Zooniverse users via email. Subsequent spikes are likely to reflect outreach efforts by Zooniverse organizers in the form of email newsletters or (more recently) through social media. In addition, projects see spikes in activity in response to coverage by mainstream media, websites, blogs, etc. (as also noted in refs. 14 and 15). To identify potential drivers of particular spikes seen in Fig. S4, we asked Zooniverse organizers for any pertinent information (excluding Galaxy Zoo Hubble that had no spikes, and Galaxy Zoo Supernovae with spikes that are not informative). We were able to obtain information on likely drivers of spikes in four of the projects. Although only suggestive, this information points to the importance of both outreach efforts via newsletters sent by Zooniverse, as well as attention from certain websites or media outlets:

- Moon Zoo spike around day 14: Newsletter to Zooniverse users
- Moon Zoo spike around day 117: Heavy traffic originating from news.bbc.co.uk
- Moon Zoo spike around day 129: Heavy traffic from cosmiclog.msnbc.msn.com
- Moon Zoo spike around day 171: Newsletter to Zooniverse users
- Old Weather spike around day 16: Newsletter to Zooniverse users
- Milkyway Project spike around day 32: Heavy traffic from sciencefriday.com
- Milkyway Project spike around day 46: Newsletter to Zooniverse users
- Milkyway Project spike around day 112: Newsletter to Milkyway Project users
- Planet Hunters spike around day 120: Heavy traffic from time.com and news.yahoo.com

Activity on weekdays vs. weekends. To explore whether participation differs between weekdays (when users may have less free time due to their regular jobs) vs. weekends (when they may have more free time), we coded each day as either falling on a weekend or being a weekday. Using this coding we compared the average daily number of hours received by each project. We find that the contributions are distributed very evenly across days of the week; the ratio of contributions received on a typical weekend day vs. weekday is 0.97 (Table S2). This analysis is only suggestive, however, because of two limitations. First, although prior work suggests that most Zooniverse users reside in the United States and the United Kingdom (16), we do not know individual users' time zones and cannot distinguish weekdays and weekends exactly (our analysis uses US central time). Second, given the high volume of contributions in the first days of a project and the important role of spikes that are likely triggered by external events, weekend vs. weekday activity may partly reflect when projects were launched and when external events occurred rather than when users prefer to contribute. Thus, although our data suggest no difference in effort levels on weekends vs. weekdays, future research is needed to validate this result.

Differences across projects. Although the qualitative patterns we find are remarkably consistent across projects, more detailed comparisons also show interesting heterogeneity among projects. For example, there is a considerable range in the number of total users (ranging from 3,186 to 28,828), the share of users who return (from 17% to 40%), the average duration of daily effort on an active day (from 7.18 to 26.23 min), and the inequality in contributions (Gini coefficient ranging from 0.77 to 0.91). This heterogeneity remains large even if we exclude the smallest project (Galaxy Zoo Supernovae), which did not operate continuously (see above). Zooniverse leaders are well aware of these differences across projects, although their explanations are largely conjectural. When asked about the large number of users in Planet Hunters, for example, one organizer suggested that users may be attracted by the opportunity to discover a completely new planet. Such discoveries are not highlighted as potential outcomes in other Zooniverse projects, where contributions consist primarily of more standardized data-related tasks (although discoveries have happened even in those contexts; see refs. 17 and 18). When asked about the high share of users that return and high levels of daily effort in the project Old Weather, several organizers noted their impression that this project has fewer casual users and a more dedicated set of core users than other projects. However, it was not clear what particular aspect of Old Weather might be responsible for attracting such a dedicated user base.

Although we will not be able to answer the question of which project features are responsible for differences in contribution patterns, we can explore and illustrate some avenues for addressing it. As noted in Table S1, projects differ with respect to a number of dimensions such as their scientific field, the type of raw data used (photographs, video, graphs visualizing numerical information, etc.), and the task users perform. Unfortunately, a sample of seven projects does not allow us to formally analyze the relationships between project characteristics and contribution patterns because projects differ across multiple dimensions simultaneously and we cannot focus on one while controlling for the others. As a first exploratory step, however, we consider the time it takes to perform one classification as a project characteristic that is both theoretically relevant and meaningful to compare across our seven projects. For example, it is conceivable that tasks that take more time to complete deter potential users that are not very interested in a project such that those who decide to contribute are less likely to drop out and may exert higher levels of effort than users in projects with lower "entry barriers." Similarly, tasks that require different amounts of time may benefit to different degrees from individuals' learning over time. To explore these possibilities, Fig. S7 plots the average time per classification in a project (x axis) against four project-level outcomes discussed earlier: the share of users who return (Fig. S7*A*), the average total effort contributed by those contributors who returned to the project (Fig. S7*B*), the Gini coefficient of the distribution of total classifications (Fig. S7*C*), and the improvement in classification speed for those top contributors who are observed for at least 7 active days (Fig. S7*D*). Fig. S7 *A* and *B* indeed suggests a positive relationship between the average time per classification on the one hand, and return rates and the average daily effort on the other. In particular, the project Old Weather has by far the longest time per classification while also having the highest rate of return users and the longest daily time spent by those return users. At the other end of the spectrum, Galaxy Zoo Supernovae has the shortest time per classification, the lowest rate of return users, and the lowest time spent per return users on an active day. The patterns for the other five projects are more ambiguous, however, with no clear relationship among the focal variables. Fig. S7*C* shows no systematic relationship between the time per classification and the degree to which total classifications are concentrated among contributors (Gini coefficient). Fig. S7*D* shows a weak negative relationship between time per classification and speed improvement,

although this seems primarily driven by Old Weather with a long time per classification but low rate of speed improvement.

Given that a sample of seven projects does not allow us to control for other project characteristics, this analysis of the relationship between the time it takes to complete one classification and project level contribution patterns should be considered only as an illustration. However, we hope that it stimulates future research on project characteristics and their relationships with project success. On the conceptual side, work is needed that complements existing descriptive studies by theorizing which features are likely to matter and why (6, 19). Empirically, large-scale data sets that include sufficient variability in project-level characteristics and project outcomes would be particularly useful. Experimental approaches that systematically manipulate key project features may be informative as well. The resulting insights could help project organizers in their efforts to attract more users and sustain their interest over time and may strengthen crowd science as an organizational mechanism to produce scientific knowledge.

1. Bianchetti RA (2014) *Looking Back to Inform the Future: The Role of Cognition in Forest Disturbance Characterization From Remote Sensing Imagery* (Pennsylvania State Univ, State College, PA).
2. Eveleigh A, Lynn S, Cox A, Jennett C (2013) I want to be a captain! Gamification in the Old Weather citizen science project. *Gamification '13: First International Conference on Gameful Design, Research, and Applications* (ACM, New York).
3. Stephan P (2012) *How Economics Shapes Science* (Harvard Univ Press, Cambridge, MA).
4. Hackman JR, Oldham GR (1976) Motivation through the design of work: Test of a theory. *Organ Behav Hum Perform* 16(2):250.
5. Brabham DC (2008) Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence (London)* 14(1):75–90.
6. Jackson C, Østerlund C, Crowston K, Mugar G, Hassman KD (2014) Motivations for sustained participation in Citizen Science: Case studies on the role of talk. *17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. Available at crowston.syr.edu/sites/crowston.syr.edu/files/MotivationinTalk(A).pdf. Accessed December 19, 2014.
7. Mason W, Watts DJ (2010) Financial incentives and the performance of crowds. *ACM SigKDD Explorations Newsl* 11(2):100–108.
8. Chandler D, Kapelner A (2013) Breaking monotony with meaning: Motivation in crowdsourcing markets. *J Econ Behav Organ* 90:123–133.
9. Franzoni C, Sauermann H (2014) Crowd Science: The organization of scientific research in open collaborative projects. *Res Policy* 43(1):1–20.
10. Silvia PJ (2006) *Exploring the Psychology of Interest* (Oxford Univ Press, Oxford, UK).
11. Krapp A, Prenzel M (2011) Research on interest in science: Theories, methods, and findings. *Int J Sci Educ* 33(1):27–50.
12. Rogers EM (2003) *Diffusion of Innovations* (Free Press, New York), 5th Ed.
13. Mahajan V, Muller E, Bass FM (1990) New product diffusion models in marketing: A review and directions for research. *Journal of Marketing* 54(1):1–26.
14. Cooper S, et al. (2010) Predicting protein structures with a multiplayer online game. *Nature* 466(7307):756–760.
15. Kwak D, et al. (2013) Open-Phylo: A customizable crowd-computing platform for multiple sequence alignment. *Genome Biol* 14(10):R116.
16. Raddick MJ, et al. (2013) Galaxy Zoo: Motivations of citizen scientists. *Astron Educ Rev* 12(1).
17. Lintott CJ, et al. (2009) Galaxy Zoo: 'Hanny's Voorwerp', a quasar light echo? *Mon Not R Astron Soc* 399(1):129–140.
18. Cardamone C, et al. (2009) Galaxy Zoo Green Peas: Discovery of a class of compact extremely star-forming galaxies. *Mon Not R Astron Soc* 399(3):1191–1205.
19. Prestopnik NR, Crowston K (2012) Citizen science system assemblages: understanding the technologies that support crowdsourced science. *Proceedings of the 2012 iConference* (ACM, New York), pp 168–176.

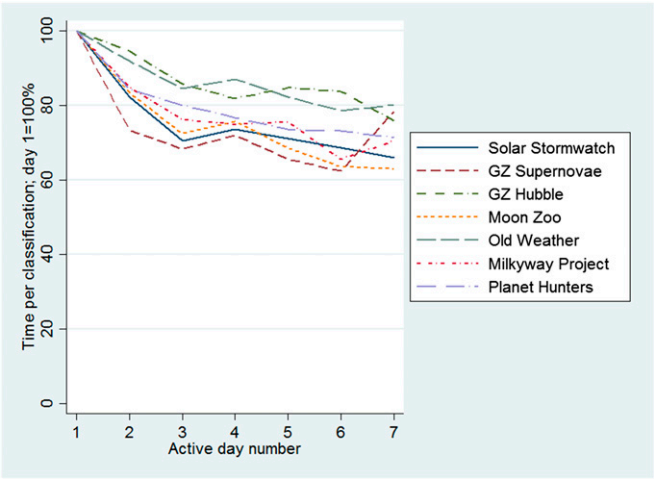

Fig. S1. Change of individuals' average time per classification over the first 7 active days in a project. Sample limited to top contributors with at least 7 active days. For comparability across projects, we set the time per classification in day 1 to 100% and show the time per classification in subsequent days relative to this reference point.

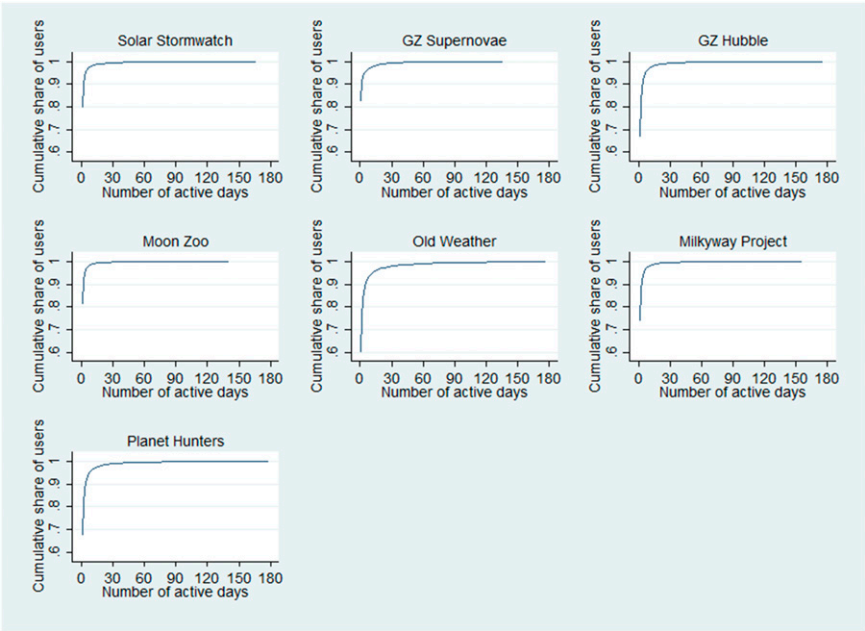

Fig. S2. Cumulative distributions of active days per user. Each panel shows the share of users (*y* axis) that has no more than a given number of active days (*x* axis) in the project. The first data point for each project (i.e., number of active days = 1) indicates the share of users who visited a project only for one day (i.e., the share of nonreturning users).

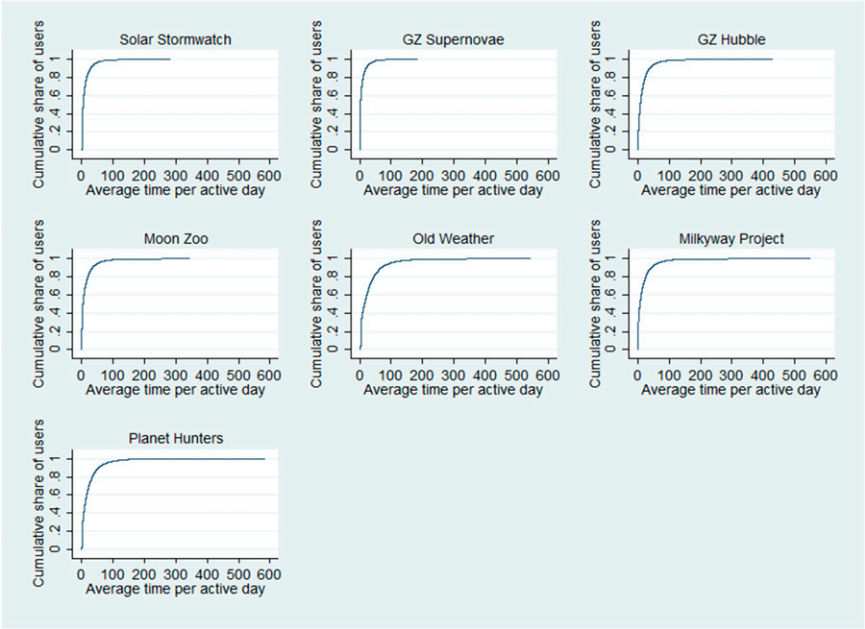

Fig. S3. Cumulative distribution of average time spent working per active day (in minutes). Average time spent working per active day is obtained by dividing an individual's total time spent on the project by the number of active days. Each panel shows the share of users (*y* axis) that spends no more than a given number of minutes (*x* axis) on the project per average active day.

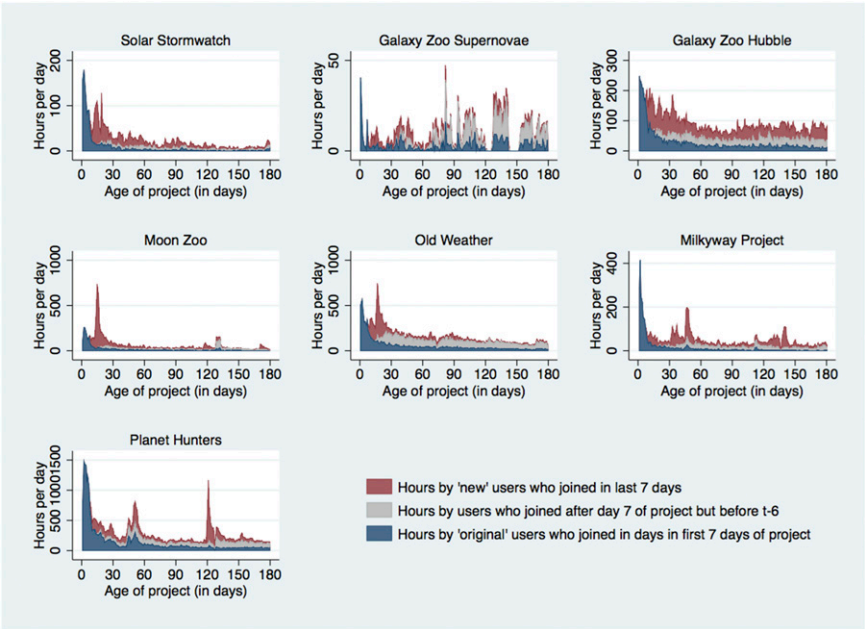

Fig. S4. Area chart of daily hours of effort received by projects during the first 180 d of each project's life. Total effort divided into hours spent by original users (users who joined in the first 7 d of the project; *Bottom*), new users (users who joined in a rolling window of the last 7 d; *Top*), and the residual group (users who joined after day 7 of the project but >6 d before the observation day; *Middle*).

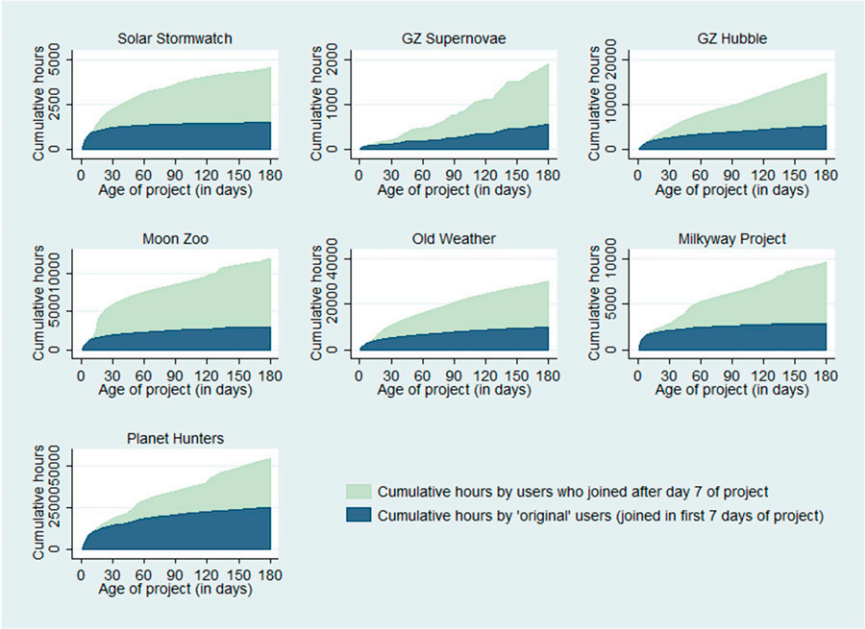

Fig. S5. Area chart showing the cumulative number of hours contributed by original users (users who joined in the first 7 d of the project; *Lower*) and all other users (*Upper*).

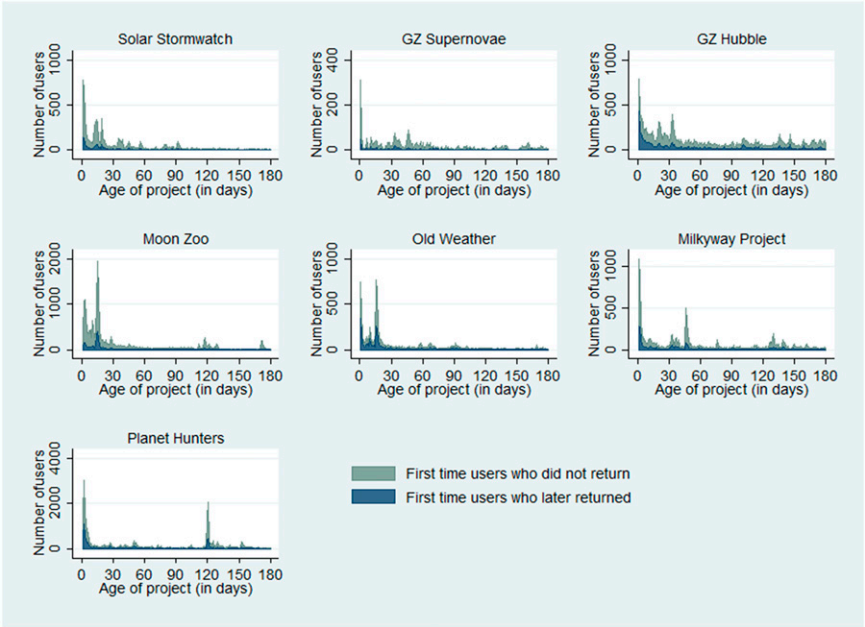

Fig. S6. Area chart showing the number of users joining a project on a particular day (first time users). Total number of first time users divided into users who returned at least once during the observation window (*Lower*) and users who did not return during the observation window (*Upper*).

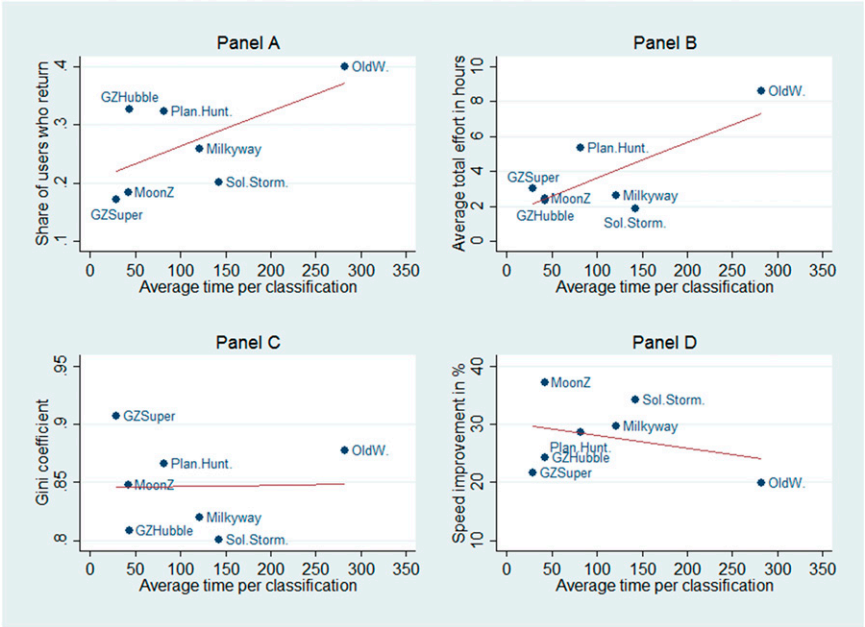

Fig. S7. Relationships between average task duration and four project level outcomes. (*A*) Share of users that return. (*B*) Average total effort of returning users. (*C*) Gini coefficient of the distribution of total classifications. (*D*) Increase in speed (based on time per classification) from active day 1 to active day 7 for top contributors with at least 7 active days. All panels include a linear fit line ($n = 7$).

Table S1. Overview of project characteristics

Characteristic	Solar Stormwatch	Galaxy Zoo Supernovae	Galaxy Zoo Hubble	Moon Zoo	Old Weather	Milkyway Project	Planet Hunters
Field	Astronomy	Astronomy	Astronomy	Astronomy	Climatology	Astronomy	Astronomy
Project goal	Generate records of solar storm occurrence and characteristics	Identify supernovae	Classify galaxies based on morphology	Build a detailed map of the surface of the moon	Generate historical climatic records	Build a detailed map of the milkyway	Identify unknown planets
Description of activity	Watch videos of solar activity and report inception and length of solar explosions	Inspect a sequence of three images and provide classifications	Inspect images of galaxies and report morphological features like shape, axes, bulge	Inspect satellite images of the Moon and report morphological features, like craters, mounds, boulders	View images of log books of historic ships and transcribe climate and location reports, fuel consumption and events	Inspect telescope images, identify certain recurring or rare features like bubbles, star clusters, EGOs and distant galaxies	Inspect starlight curves registered by the Kepler spacecraft and report possible planet transits
Raw data shown per single submission	Paired B/W videos of 1 min of solar activity taken simultaneously from two spacecrafts (Stereo Ahead and Stereo Behind)	Three B/W low-resolution images of stars, of which the third image is obtained as a subtraction from the prior two	One colored image of distant space taken from the NASA's Hubble Space Telescope	One B/W image of moon surface taken from NASA's Lunar Reconnaissance Orbiter	One colored lens-magnified portion of an image of a logbook page from a navy ship	One colored image of a portion of the Milkyway taken from the Spitzer Space Telescope	One B/W chart report of star brightness during a quarter
Activity type	Categorization and video tagging	Categorization	Categorization	Image tagging	Transcription from image	Image tagging	Categorization and image tagging
Cognitive tasks involved	Judgment	Classification, judgment	Classification, comparison, judgment	Comparison, classification	Deciphering, judgment	Comparison, classification, judgment	Comparison, judgment
Common disturbance	Solar wind, chunky image-stills	Stars overlap, low resolution, halos, artifacts	Stars, artifacts, color shades, low-resolution	Halos, absence of perspective, blurred regions	Paper deterioration, unreadable handwriting, ink spots, spelling mistakes	Poor color resolution, small-scale objects, image artifacts	Gaps (e.g., downtime of telescope), irregular/fuzzy patterns
Project start	23-Feb-10	29-Mar-10	23-Apr-10	11-May-10	13-Oct-10	7-Dec-10	16-Dec-10
Observation end date	21-Aug-10	24-Sep-10	19-Oct-10	6-Nov-10	10-Apr-11	4-Jun-11	13-Jun-11

Table S2. Key statistics for all seven projects, averages across projects and totals (sums over projects)

Statistic	Sol.Storm	GZSuper	GZHubble	MoonZ.	OldW.	Milkyway	Plan.Hunt.	Average	Total all projects
Total hours and classifications, monetary valuation									
Total users	9,151	3,186	18,837	20,614	8,291	11,479	28,828	14,341	100,386
Total hours contributed	4,566	1,893	16,923	11,985	30,050	9,611	54,511	18,506	129,540
FTE (8 h/d, 5 d/wk, for 180 d)	4.44	1.84	16.45	11.65	29.22	9.34	53.00	17.99	125.94
Counterfactual cost at $12/h	54,789	22,717	203,080	143,817	360,604	115,337	654,130	222,068	1,554,474
Total classifications	108,660	284,137	1,869,323	1,954,661	416,751	302,797	3,130,672	1,152,429	8,067,001
AMT cost per classification ($)	1.000	0.105	0.140	0.175	0.650	0.150	0.155	0.339	
Counterfactual cost on AMT ($)	108,660	29,834	261,705	342,066	270,888	45,420	485,254	220,547	1,543,827
Distribution of effort and classifications									
Total classifications by top 10	77,421	250,601	1,383,808	1,613,984	349,623	217,706	2,594,718	926,837	
Share of total classifications by top 10	0.71	0.88	0.74	0.83	0.84	0.72	0.83	0.79	
Gini coefficient for total classifications	0.77	0.91	0.82	0.88	0.88	0.80	0.88	0.85	
Total hours by top 10	3,198	1,608	11,549	8,318	24,620	6,580	42,586	14,066	
Share of total hours by top 10	0.70	0.85	0.68	0.69	0.82	0.68	0.78	0.74	
Gini coefficient for hours	0.80	0.91	0.81	0.85	0.88	0.82	0.87	0.85	
Average time per classification (in seconds)	142.80	28.70	42.97	41.86	282.25	121.47	82.23	106.04	
Speed advantage of top contributors									
Average time per classification by top 10 contributors	158.44	25.64	31.67	25.99	270.45	120.13	64.96	99.61	
Average time per classification by non-top 10	141.13	29.04	44.23	43.61	283.55	121.61	84.14	106.76	
Speed advantage top 10	−12%	12%	28%	40%	5%	1%	23%	14%	
Average time per classification on first day (in seconds)	146.30	29.85	45.03	45.09	296.02	128.01	87.31	111.09	
Average time per classification by top 10 on first day	187.82	34.10	41.07	44.05	345.36	155.16	86.14	127.67	
Average time per classification by non-top 10 on first day	141.86	29.38	45.47	45.21	290.58	125.03	87.44	109.28	
Speed advantage of top 10 on first day	−32%	−16%	10%	3%	−19%	−24%	1%	−11%	
Speed increase of top 10 contributors over first 7 active days	34%	22%	24%	37%	20%	30%	29%	28%	
Individual dynamics									
Share of users who return at least once	0.20	0.17	0.33	0.18	0.40	0.26	0.32	0.27	
Share of contributions by returning users	0.76	0.88	0.83	0.80	0.95	0.78	0.92	0.85	
Average duration of daily effort	10.24	7.18	15.51	12.34	26.23	15.81	21.23	15.51	
90th percentile of average duration of daily effort	28.20	20.19	37.65	33.98	68.02	41.97	53.90	40.56	
Average total effort per user (in hours)	0.50	0.59	0.90	0.58	3.62	0.84	1.89	1.28	
Average total effort per returning user (in hours)	1.89	3.00	2.31	2.46	8.60	2.63	5.34	3.75	
Break between active days 1 and 2 (in days)	3.27	12.27	5.81	5.79	2.40	3.43	3.63	5.23	
Break between active days 6 and 7 (in days)	6.53	9.28	9.45	10.82	6.36	8.18	7.50	8.30	
Project level dynamics									
Share of "original" users (joined in first 7 d)	0.26	0.15	0.14	0.22	0.20	0.24	0.31	0.22	
Share of total hours contributed by "original" users	0.34	0.30	0.31	0.25	0.33	0.30	0.46	0.33	
Number of spikes in activity	7	28	0	6	1	5	3	7.14	
Ratio of hours received on avg. weekend day vs. weekday	0.98	0.89	0.90	0.94	1.02	1.00	1.05	0.97	

Statistics are ordered according to the flow of the discussion in the main text.

Table S3. Changes in classification speed over time

Variable	1 Sol.Storm	2 GZSuper	3 GZHubble	4 MoonZ.	5 OldW.	6 Milkyway	7 Plan.Hunt.
Active day 1	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Active day 2	−38.342*	−11.564	−2.928	−12.978*	−29.852*	−31.822*	−16.031*
	[8.192]	[6.012]	[1.835]	[3.541]	[7.378]	[7.594]	[1.992]
Active day 3	−64.148*	−13.798†	−7.678*	−21.624*	−57.020*	−50.306*	−20.350*
	[7.778]	[6.359]	[1.926]	[3.956]	[7.544]	[8.015]	[2.083]
Active day 4	−57.245*	−12.187	−9.808*	−19.073*	−47.815*	−52.887*	−23.703*
	[8.602]	[6.586]	[1.976]	[3.933]	[7.637]	[7.722]	[2.253]
Active day 5	−62.703*	−14.946†	−8.254*	−24.603*	−65.477*	−51.673*	−26.983*
	[9.483]	[6.582]	[2.266]	[3.725]	[7.273]	[8.408]	[2.237]
Active day 6	−68.103*	−16.363†	−8.743*	−28.591*	−79.168*	−72.611*	−27.404*
	[8.621]	[6.487]	[2.378]	[3.768]	[7.488]	[7.974]	[2.206]
Active day 7	−74.100*	−9.371	−13.024*	−29.044*	−73.485*	−62.353*	−29.164*
	[8.984]	[7.507]	[2.179]	[3.803]	[7.897]	[7.519]	[2.197]
Constant	217.052*	43.423*	53.932*	78.284*	368.841*	210.673*	101.907*
	[6.104]	[5.220]	[1.434]	[2.724]	[5.304]	[5.576]	[1.548]
Observations	1,932	910	5,299	2,863	4,242	2,471	10,864

Regressions of the average time taken per classification on an active day, using OLS with individual fixed effects. Sample restricted to top contributors with at least 7 active days. Analysis at the level of the person-day, limited to the first 7 active days (7 observations for each user). Standard errors in brackets.

*Significant at 1%.

†Significant at 5%.

Table S4. Regressions predicting whether a user returns to a project, using OLS

	Pooled		Solar Stormwatch		GZ Supernovae		GZ Hubble		Moon Zoo		Old Weather		Milkyway Project		Planet Hunters	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Variable	returned30	returned30	returned30	returned30	returned30	returned30	returned30	returned30	returned30	returned30	returned30	returned30	returned30	returned30	returned30	returned30
Start day		−0.0002	0.0003*		0.0005†		−0.0003†		0.0004†		−0.0009†		0.0001		−0.0004**	
		[0.0001]	[0.0001]		[0.0002]		[0.0001]		[0.0001]		[0.0001]		[0.0001]		[0.0001]	
Original user	0.0223			−0.0099		−0.032		0.1631†		−0.0273†		0.0736†		0.0159		−0.0035
	[0.0252]			[0.0094]		[0.0163]		[0.0103]		[0.0060]		[0.0138]		[0.0096]		[0.0062]
Start during	−0.0139	−0.0158	0.0158	0.0143					0.0384†	0.0304†	−0.0997†	−0.0620†	−0.0487†	−0.0423†	−0.0299**	−0.0626**
spike	[0.0287]	[0.0230]	[0.0192]	[0.0195]					[0.0071]	[0.0073]	[0.0180]	[0.0180]	[0.0138]	[0.0141]	[0.0088]	[0.0081]
GZ Supernovae	−0.0480†	−0.0479†														
	[0.0032]	[0.0014]														
GZ Hubble	0.1012†	0.1024†														
	[0.0034]	[0.0023]														
Moon Zoo	−0.0337†	−0.0349†														
	[0.0036]	[0.0027]														
Old Weather	0.1944†	0.1936†														
	[0.0018]	[0.0012]														
Milkyway	0.0439†	0.0468†														
Project	[0.0009]	[0.0029]														
Planet Hunters	0.1012†	0.1066†														
	[0.0030]	[0.0045]														
Constant	0.1840†	0.1965†	0.1784†	0.1913†	0.1129†	0.1450†	0.3043†	0.2630†	0.1337†	0.1536†	0.4240†	0.3722†	0.2308†	0.2320†	0.3182**	0.3004**
	[0.0072]	[0.0058]	[0.0057]	[0.0051]	[0.0101]	[0.0072]	[0.0056]	[0.0037]	[0.0035]	[0.0033]	[0.0082]	[0.0065]	[0.0061]	[0.0050]	[0.0042]	[0.0039]
Observations	93,610	93,610	8,840	8,840	2,842	2,842	16,684	16,684	19,870	19,870	7,900	7,900	10,687	10,687	26,787	26,787
R^2	0.0273	0.0272	0.0009	0.0002	0.0038	0.0012	0.0008	0.0172	0.0031	0.0025	0.007	0.0061	0.0011	0.0012	0.0039	0.0021

Models address censoring by limiting the sample to users who are observed for at least 30 d and by coding the dummy dependent variable as 1 if a user returned to the project within those 30 d. Standard errors in brackets.

*Significant at 5%.

†Significant at 1%.