

# Intervention Strategies for Increasing Engagement in Crowdsourcing: Platform, Predictions, and Experiments

Avi Segal and Ya'akov (Kobi) Gal  
Ben-Gurion University of the Negev, Israel

Ece Kamar and Eric Horvitz  
Microsoft Research, Redmond WA

Alex Bowyer and Grant Miller  
University of Oxford, U.K.

## Abstract

Volunteer-based crowdsourcing depend critically on maintaining the engagement of participants. We explore a methodology for extending engagement in citizen science by combining machine learning with intervention design. We first present a platform for using real-time predictions about forthcoming disengagement to guide interventions. Then we discuss a set of experiments with delivering different messages to users based on the proximity to the predicted time of disengagement. The messages address motivational factors that were found in prior studies to influence users' engagements. We evaluate this approach on Galaxy Zoo, one of the largest citizen science application on the web, where we traced the behavior and contributions of thousands of users who received intervention messages over a period of a few months. We found sensitivity of the amount of user contributions to both the timing and nature of the message. Specifically, we found that a message emphasizing the helpfulness of individual users significantly increased users' contributions when delivered according to predicted times of disengagement, but not when delivered at random times. The influence of the message on users' contributions was more pronounced as additional user data was collected and made available to the classifier.

## 1 Introduction

Volunteer crowdsourcing projects such as citizen science efforts rely on volunteers, rather than paid workers, to perform tasks. For example, Zooniverse, the world's most popular platform for citizen science, engages millions of active volunteers who accelerate scientific discovery by analyzing data as individuals and as groups [Simpson *et al.*, 2014]. User contributions in crowdsourcing platforms typically follow a power-law distribution, where the majority of participants make very few contributions [Preece and Shneiderman, 2009]. Thus, extending and sustaining user engagement in crowdsourcing platforms is important and remains an active area of research [Eveleigh *et al.*, 2014; Segal *et al.*, 2015].

Paid crowdsourcing platforms such as Amazon mechanical Turk or CrowdFlower incentivize workers to remain active and to make high quality contributions with monetary incentives [Ipeirotis, 2010; Horton and Chilton, 2010; Tran-Thanh *et al.*, 2015]. In contrast, volunteer-based platforms rely on intrinsically motivated participants who do not get paid for their contributions. While some volunteers may become recurrent contributors, the vast majority of volunteer participants are "dabblers" [Eveleigh *et al.*, 2014], who make a small number of contributions before disengaging and never returning. Despite this casual and non-committed participation pattern, dabblers contribute a substantial amount of the overall effort in these platforms. Even a small increase in the contribution rates of these users can lead to a significant improvement in productivity for the platform.

We report on a study of intervention strategies aimed at increasing engagement in volunteer-based crowdsourcing. We employ machine learning to build predictive models that can be used in real time to identify users who are at risk for disengaging soon from the system. We describe a real-time platform that uses predictions of disengagement to guide the delivery of messages designed to increase users' motivation to contribute. The approach addresses two key challenges of intervention design: First, we evaluate different intervention strategies with respect to balancing the potential disruption of delivering messages with the benefits of the expected improvement in users' productivity. Second, we explore the use of predictive models to identify the best time to intervene. Intervening too early may not address the loss of motivation that precedes disengagement, while intervening too late may miss participants who disengage before this time.

We evaluated our approach on Galaxy Zoo ([www.galaxyzoo.org](http://www.galaxyzoo.org)), one of the largest citizen science projects on the web, in which volunteers provide classifications on celestial bodies. Galaxy Zoo has attracted over 120,000 users who have classified over 20 million galaxies since the inception of the project's fourth version in 2012. However, over 30% of users complete fewer than 10 classifications (See Figure 1) and most users do not come back for a second session.

We used three types of motivational messages that were designed to address different motivational issues identified in citizen science by prior work. These messages emphasized how users contribute to the project, their sense of community, or tolerance of the project to individuals' potential mistakes.

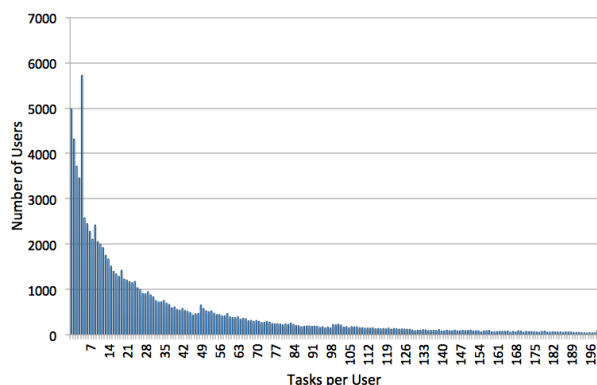


Figure 1: Distribution over numbers of classifications per user in Galaxy Zoo.

We compared the effects of these different messages on users' behavior in the system when guided by policies that use predictions about forthcoming disengagement versus when delivered at random times. Our experiments were designed collaboratively with Galaxy Zoo leads and administrators.

We found that the message emphasizing the value of users' individual contributions to the platform when timed based on disengagement predictions was able to significantly increase the contribution of users without decreasing the quality of their work or increasing the amount of time spent in the system. Using the same message, but displaying this message at random times, was not effective. A follow-up study compared the effects of using different thresholds on the likelihood of forthcoming disengagement of users to initiate intervention messages, which consequently influenced when users received interventions.

We make three contributions: First, we provide a methodology and platform for investigating and controlling influences on engagement in crowdsourcing aimed at extending engagement via coupling machine learning and inference with intervention design. Second, we report on a set of experiments with messaging content and timing that explores the efficacy of alternate strategies when applied to a live volunteer-based crowdsourcing system. Third, we show that it is important to consider both message content and the timing of the delivery of messages. Code, data and accompanying information to this paper can be found at <http://tinyurl.com/ztujcvz>.

## 2 Related Work

Multiple studies have documented the long-tail distribution of users' behavior in volunteer-based platforms such as Wikipedia [Preece and Shneiderman, 2009] and citizen science [Eveleigh *et al.*, 2014; Sauerermann and Franzoni, 2015].

Mao *et al.* [2013] developed a predictor for disengagement in volunteer-based crowdsourcing. They considered 150 features including statistics about volunteers' characteristics, the tasks they solved, and their history of prior sessions on the system. They demonstrated the effects of different session lengths and window sizes on the prediction accuracy. Their model was tested in an offline setting with holdout data, as-

suming multiple session histories are available for each user. Inspired by their call for real-time use of their approach, we extended earlier work in two ways: First, we adopted their model to a real-time setting in which users may have limited amount of history (or none at all), and selected a minimal number of features that provide reasonable predictor performance. Second, we implemented the predictive model and used it to guide a set of real-time, live interventions with users engaging with the system.

The study of intervention mechanisms for improving users' productivity is receiving increasing attention from the social and computational sciences. Some works have focused on increasing users' intrinsic motivation by generating messages to users, whether by framing a task as helping others [Rogstadius *et al.*, 2011], reminding users of their uniqueness [Ling *et al.*, 2005] or making a direct call for action [Savage *et al.*, 2015]. Anderson *et al.* [2013] developed a model for the influence of merit badges in the stack-overflow platform. They showed how the community behavior changes once users get closer to the badges frontier, and gave insights on the optimal placement of badges in such a system. In another work, Anderson *et al.* [2014] implemented a large-scale deployment of badges as incentives for engagement in a MOOC. They found that making badges more salient produced an increase in forum engagement. Segal *et al.* [2015] demonstrated that they could increase the return rate of volunteers to a citizen science system by sending motivational emails several days after they stopped making contributions.

Lastly, we review work on modeling and managing interruptions associated with notifications. Horvitz and Apacible [2003] used machine learning with Bayesian structure search to build models that infer the cost of interrupting users over time considering their interactions and tasks on computing devices, visual and acoustical analyses, and data drawn from online calendars. Shrot *et al.* [2014] used collaborative filtering approaches to predict the cost of interruption by exploiting the similarities between users and used this model to inform an interruption management algorithm. Kamar *et al.* [2013] showed that modeling interruptions as a planning under uncertainty problem can improve agents' performance while interacting with people.

## 3 Problem and Approach

The problem of if, when, and how to generate interventions in order to increase users' engagement in volunteer-based crowdsourcing can be formalized as a sequential decision-making problem under uncertainty. With an intervention strategy centering on the use of real-time messaging, we need to balance the potential distraction of presenting an intervention message to the user with the expected short and long-term benefit of the message on their contributions. The state in the problem formalization is a tuple that represents a user's interactions with the platform in current and prior sessions, including information about past interventions that were administered for this user. The action set defines a rich design space of intervention strategies: when to intervene (e.g., periodically, when a likelihood level of disengagement is reached, etc.); how to channel the intervention (e.g., email, pop-up,

etc.); how the intervention is administered (e.g., text, audio, graphical, video, mixed, etc.); the content of the intervention (e.g., motivational message, performance feedback, community message, etc.), and its duration (e.g., for a given time period, until the user has acknowledged the intervention, etc.).

Formulating, parameterizing, and solving this decision-making problem to identify ideal intervention policies introduces a number of challenges. First, learning about the outcomes of different interventional policies would entail studying and probing volunteer crowdworkers over a large space of state and action combinations. Second, we do not have an understanding of the influence of interventions on contributions. Thus, it is difficult to formulate an objective function for use in formal sequential decision making analysis. For example, a focus on optimizing the number of tasks completed by users may lead to a large number of generated interventions but ultimately reduce the quality of users' contributions.

Our approach is to simplify the general decision-making problem into a simpler decision problem: we use predictions about forthcoming disengagement to control the timing of the display of different messages. Specifically, this approach consists of the following three steps: First, we employ supervised learning to build a model that is used in real-time to predict whether a user is about to disengage from the system within a given time window. The learned model uses task-independent features relating to users' past and present interactions with the system. Second, we employ an intervention strategy that targets users whose predicted likelihood of disengagement within the window exceeds a given threshold. The intervention consists of different messages designed to address the motivational issues that have been identified in prior studies [Raddick *et al.*, 2010; Reed *et al.*, 2013; Eveleigh *et al.*, 2014; Segal *et al.*, 2015] as reasons that users disengage from the system. Third, we perform a controlled study to evaluate different intervention strategies in a live citizen science platform.

We used three types of motivational messages aimed at emphasizing: (1) the contributions of users to the system, (2) users' sense of community, and (3) the tolerance of the project to an individual's potential mistakes. We compared the effects of these messages on users' behavior in the system when guided by policies that use predictions about forthcoming disengagement versus when delivered at random times.

We hypothesized the following: (1) Users in the prediction based intervention groups would be more productive (as measured by contributions and time in the system) than with the random intervention conditions (and a control group receiving no interventions), without harming the quality of their contributions. (2) The influence of the intervention on users' productivity depends on the message content, the timing of the message, and the confidence threshold on likelihood of forthcoming disengagement.

## 4 Predicting Disengagement

We now describe the first step of our approach, where a predictive model is used to identify users that are at risk of disengaging from the platform. We follow the formalization of this prediction problem by Mao *et al.* [2013]. We assume a gen-

eral crowdsourcing platform in which users work individually to solve tasks. A session of a user is defined as a contiguous period of time that the user is engaged with the platform and is completing tasks. The prediction problem is a binary classification problem; given the current state of a user (encapsulating both user history and current session), predict whether the user will disengage and end the current session within a given time window. We consider all tasks which are 30 minutes or less apart from each other as belonging to the same session of activity. We use 5 minutes as the disengagement time window in our prediction problem.

Our dataset included 1,000,000 classifications from 13,475 unique users who were active in Galaxy Zoo in the 12 months preceding the experiment. To illustrate the casual participation behavior in our population, the average number of sessions per user was 2.6, and 67% of the users stayed for only one session. An instance to be used in learning or inference is comprised of features describing the user's past interactions up to the present task, and a binary label determining whether the user will disengage from the system within the 5 minute time window. From the above dataset we used 500,000 instances for training, 250,000 instances for validation and 250,000 instances for testing. The instances in the validation and test sets followed those in the training set in chronological order, preserving temporal consistency.

In training our model, we used the 16 most informative featured identified by Mao *et al.* We confirmed that using this smaller feature set did not significantly affect the predictive performance when compared to the results they reported. The top five features included in the predictive model are, presented in decreasing order of importance: user's average session time over all sessions so far; user's average dwell time in current session (measured in the number of seconds between two consecutive tasks); user's session count; number of seconds elapsed in current session; difference between the number of tasks performed by user in the current session and median number of tasks in previous 10 sessions of this user (or null if user has completed fewer than 10 sessions).

Figure 2 (top) shows the predictive performance of the classifier in terms of area under the receiver-operator characteristic curve (AUC), as a function of the number of session histories available for users. As shown in the figure, the trained classifier performs better than a baseline based on selecting the most likely class (AUC of .5) and the performance increases proportionally to the amount of historical data available for individual users.

Figure 2 (bottom) shows the effect of different disengagement thresholds on the predicted performance, measured in terms of precision (percentage of correctly classified dropouts out of all predicted dropouts), recall (percentage of correctly classified dropouts out of all relevant dropouts), and accuracy (percentage of correctly classified dropouts and continued engagements out of all predicted dropout and continuations). As displayed in the figure, increasing the confidence threshold also increases the precision, because the predictor becomes more conservative about deciding whether a user is at risk for disengagement within the target window. However, this raised threshold also reduces recall, because a conservative strategy will miss relevant dropouts. The accuracy

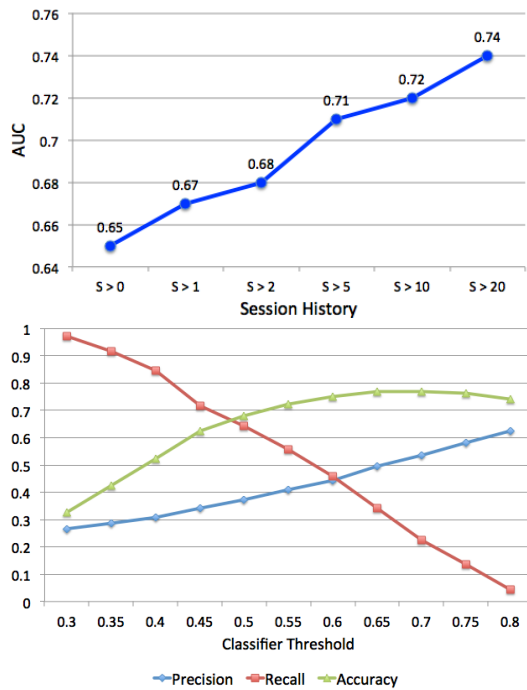


Figure 2: Performance in predicting disengagement. In top figure,  $S > x$  means number of sessions is larger than  $x$ .

of the predictor steadily increases as the confidence threshold grows, before leveling out around 0.75. Based on these results, we selected a confidence threshold of 0.5 to initiate interventions in our experiments, which was expected to provide a good balance between precision and recall without significantly compromising the prediction accuracy.

## 5 Designing Intervention Strategies

We now describe the second step of our approach, centering on the design of intervention messages that are promising for increasing the motivation of users in the platform. The intervention messages (shown in Table 1) were developed in accordance with the administrators of Galaxy Zoo, and directly address issues that have been shown in prior work to reduce the motivation of volunteers in citizen science. The helpfulness type message emphasized to individual users that their contributions are valuable to the Galaxy Zoo project. The community type message emphasized the collective nature of Galaxy Zoo project and its sense of community. The anxiety-reduction type message emphasized the tolerance to individual mistakes by volunteers, addressing the fear of making mistakes in classification, which has been documented in several studies on motivation in citizen science [Eveleigh *et al.*, 2014; Segal *et al.*, 2015].

We made the following decisions to minimize the disruption to participants associated with the delivered messages, in accordance with guidance from the Galaxy Zoo administrators. First, the messages were shown for 15 seconds or until the message was closed by the user. Second, we generated an intervention message *at most* once per session for each

Table 1: Intervention messages used in the first study

Type	Message	Cohort
Helpful	Please don't stop just yet. You've been extremely helpful so far. Your votes are really helping us to understand deep mysteries about galaxies.	Random-Helpful Predicted-Helpful
Community	Thousands of people are taking part in the project every month. Visit Talk at <a href="http://talk.galaxyzoo.org">talk.galaxyzoo.org</a> to discuss the images you see with them.	Random-Community Predicted-Community
Anxiety	We use statistical techniques to get the most from every answer; So, you don't need to worry about being "right". Just tell us what you see.	Random-Anxiety Predicted-Anxiety

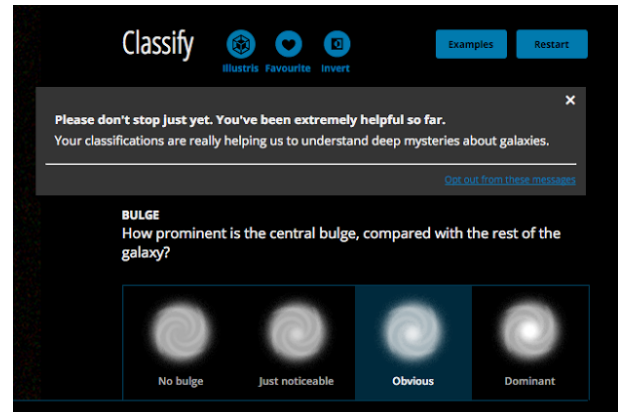


Figure 3: Intervention message overlaid on Galaxy Zoo screen (partial view).

user. Third, the intervention was introduced using a window that smoothly integrated within the Galaxy Zoo GUI using a "slide-in" animation (see Figure 3). Lastly, in accordance with the Galaxy Zoo administrations, the intervention window included an option to opt out of receiving any additional messages. The study received ethics approval from the Institutional Review Board of Ben-Gurion University.

## 6 Empirical Studies

We now describe the third step of our approach, consisting of two studies for evaluating the effect of different intervention strategies on users' behaviors in the system.

### 6.1 Effect of Timing and Message Content

In the first experiment we compared the effects of different message contents and the timing of the intervention on the contributions of users in the platform. We created two cohorts of users for each message type. For the first cohort, the timing of the message was guided using the predictive model. For the second, the timing of the message was distributed randomly. Thus, users in the Predicted-Helpful ("P-Helpful") cohort were targeted by the helpfulness intervention message when they were predicted to disengage within the target horizon by the model, and users in the Random-Helpful ("R-Helpful") cohort were targeted by the helpfulness intervention message randomly (and similarly for the Predicted-Community and Predicted-Anxiety cohorts).

The amount of time to an intervention in the random condition was determined by fitting a Poisson distribution to the

prior session history we obtained from Galaxy Zoo. The intervention time was then drawn uniformly between the limits of 0 (i.e., intervene immediately) and a session length that was sampled from this distribution. Note that in practice, the user may already have disengaged from the system by the determined intervention time. To keep the number of interventions equal for each cohort, we ran a pre-experiment simulation to compute the number of interventions that would be generated by the prediction based approach on past data and normalized the random condition to match this number of interventions. Using this strategy we were able to generate nearly equal number of interventions between the different cohorts.

A total of 3,377 users were considered in the study, of which 2,544 (75%) were new users with no prior history of interaction in the system. The study took place between August 10 and September 20, 2015. Users logging on to the system during this time period were randomly divided between the six cohorts described above and an additional cohort (the control group) which received no interventions. All cohorts included 567 volunteers each, except the Predicted-Anxiety cohort which included 565 volunteers. In total, 4,168 interventions were generated for all of the intervention cohorts. At the request of the Galaxy Zoo administrators, we left out of the study a small minority of super-users with a contribution rate that was greater than three standard deviations from the mean contribution rate for all cohorts. This sub-population included 33 users (0.1% of total participants) with an average contribution rate of 456 tasks per session (the remaining user population had less than 50 tasks per session on average). These super users were removed from the study since they had already established themselves as persistent contributors with significantly different contribution patterns and they were not a target population for our study.

Figure 4 (top) shows the average contribution rates for each of the cohorts. There was no statistically significant difference between the contribution rates of users in the random intervention cohorts and the control group. The figure also shows that the users in the P-Helpful cohort generated 19.6% more contributions than users in the R-Helpful cohort which saw the same message ( $p < 0.05$ , analysis of variance).

Figure 4 (bottom) shows the evolution of contribution rates for the seven conditions as the number of sessions for each user increases. As shown in the figure, the contribution rates for the users in the P-Helpful cohort steadily increases as the users in this cohort complete more sessions over time and engagement predictions become more informative as a result. In contrast, none of the other conditions expressed any significant increase in contribution rates. We note that, although the contribution of the P-Community cohort increased over time, this trend was not statistically significant.

Contrary to our expectations, none of the cohorts stayed longer in the system when compared to the control (871 seconds). Therefore, to explain the higher contribution rates for the users in the P-Helpful cohort, we analyzed the dwell time (the average number of seconds between task submissions) for the different intervention cohorts. We found that the dwell time following intervention for the P-Helpful cohort (26 seconds) was significantly shorter than the dwell time of the control group (33.5 seconds).

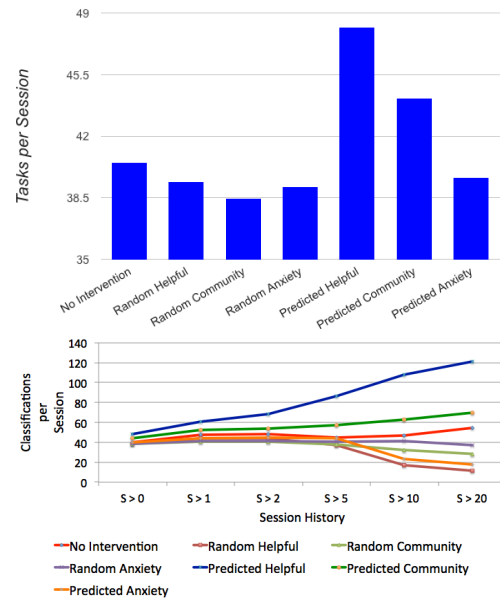


Figure 4: Comparison of number of tasks performed overall (top) and by prior sessions for all cohorts (bottom). In bottom figure,  $S > x$  means number of sessions is larger than  $x$ .

A possible consequence of the faster turnaround time for users solving tasks in this cohort is a decrease in the quality of their contributions. Since gold-standard answers to Galaxy Zoo tasks are not available, we instead used user agreement as the metric for quality. User agreement is commonly tracked as a quality metric in crowdsourcing platforms and is the basis for the aggregation algorithms such as Dawid-Skene [Ipeirotis *et al.*, 2010]. We computed the agreement score for each cohort by iterating over all Galaxy Zoo tasks worked on by the users in this cohort during the experiment. For each task, we computed the KL-divergence between the distribution of classifications collected from the cohort to the distribution of classifications collected from Galaxy Zoo since its launch in 2012 until the start of our experimentation and then averaged over all tasks. Our analysis showed no statistically significant difference between the KL divergence of the different cohorts. This analysis supports the conclusion that the quality of the work for users in the P-Helpful cohort was not different from that of the other cohorts and that the speed up from this intervention did not lead to a decrease in the quality of work.

To summarize the first study, we demonstrated that both the timing and the content of messages are important design choices in an intervention strategy aimed at improving volunteer engagement. One potential explanation of the influence of the helpful message is that it resonates with participants interest in making a socially beneficial contribution, with its emphasis of the principle of volunteerism. Nonetheless, without controlling the timing of the intervention based on predictions of forthcoming disengagement, this message is not effective (see the performance of the R-Helpful cohort). In contrast, the community type message, which encouraged users to go and explore when they are predicted to be readying



to disengage (and perhaps losing interest) may actually stimulate users to exit the task stream at focus of attention; the message is coupled with a link to a community home page. Further, we suspect that overcoming “classification anxiety” may require more intensive psychological support and modulation (e.g., through practice and reassurance on a number of occasions) than a single supportive message. Lastly, we attribute the increase in contributions over time for users in the P-Helpful cohort to the improvement in predictor performance as additional data is collected.

## 6.2 Effect of Threshold

In the second experiment we explored the sensitivity of results to changes in the probability threshold used to control the display of messages. That is, we selected different thresholds on the inferred likelihood of forthcoming disengagement required to initiate interventions. We aimed at understanding the ideal tradeoff between precision and recall in guiding the delivery of intervention messages. Based on the results from the previous study, all cohorts were targeted with the helpful intervention message based on disengagement predictions but differed in the threshold value used to predict if users would disengage within the designated time frame of 5 minutes.

Users in the low-threshold, medium-threshold and high-threshold cohorts were assigned a predictor using threshold values of 0.3, 0.5, and 0.7 respectively. These numbers were chosen because they were shown to have a significant impact on the number of interventions generated in a simulation that we ran in advance (Figure 5 top).

Figure 5 (top) shows the simulation of the number of interventions that would be generated for different threshold values and available histories of users (computed on the data collected from the control condition of experiment 1). As shown by the figure, varying the threshold used by the disengagement predictor affects how many interventions are generated for different users, depending on their past interactions. The number of generated interventions decreases for all histories as the confidence threshold of the predictor increases.

The study was conducted between December 18th, 2015 and January 5th, 2016. A total of 1,290 users participated in the study, of which 837 (65%) were new users with no prior history of interaction in the system. The low- and high-threshold cohort included 322 users each, while the medium-threshold cohort included 323 users. In total, 1,529 interventions were generated for all cohorts.

Figure 5 (bottom) shows the contribution rates for the different cohorts. As shown in the figure, the rate of contributions for the medium-threshold cohort was significantly higher than that of the other cohorts ( $p < 0.05$ , analysis of variance). This result demonstrates the effect of the prediction threshold on users’ contribution rates for interventions using the helpful-message type. While all three intervention cohorts outperformed the no-intervention cohort, the 0.5 prediction – the agnostic threshold used in the first experiment – achieved the best results.

## 7 Conclusion

We described a methodology and experiments aimed at exploring challenges and opportunities with extending the en-

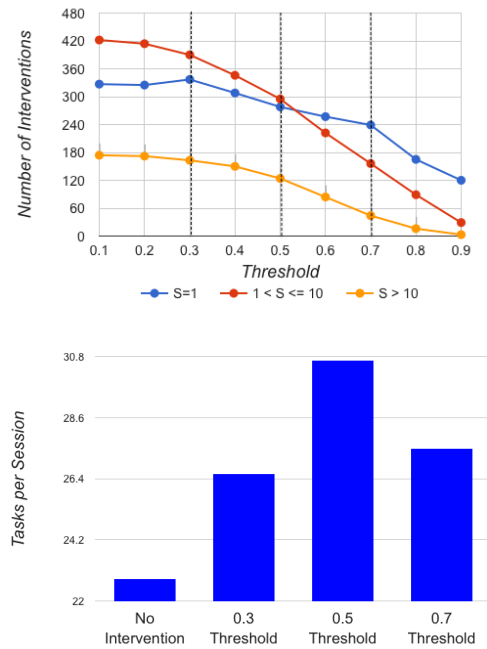


Figure 5: Effect of session history on number of interventions for different prediction thresholds (top) and effect of prediction thresholds on contributions (bottom). In top figure,  $S$  means number of sessions.

agement of users in volunteer-based crowdsourcing platforms. We focused on the Galaxy Zoo system, one of the largest crowdsourcing efforts on the web. We constructed a predictive model from a large corpora of data about Galaxy Zoo volunteers predicting when users would soon disengage from the system based on observations about their activities and their histories. We used the model to generate real-time interventions based on the inferred likelihood that volunteers engaged with the system would soon disengage. The interventions were messages designed to address recognized challenges with the motivation of volunteers assisting with citizen science problems. Our evaluations highlighted the joint effect of intervention timing and message content on the contributions of Galaxy Zoo users. We found that a message emphasizing that users individual contributions were important was able to significantly increase the contribution of users without decreasing the quality of their work, but that this messaging intervention had a significant effect only when guided by the predictive model.

We see several directions for future work. We wish to further investigate the influences and interactions between message content and timing so as to better understand the role of each factor and their combined effects on crowdworkers. We wish to study other sets of interventions, including the modulation of task type, changing difficulty of tasks, and such factors as the aesthetics, interestingness, and inspirational influence based on visual or other properties of tasks. We also see opportunity to extend this work to compute intervention policies that consider richer multi-step plan-based methodologies and designs that recognize different types of users and that

personalize the messages to different cohorts. Finally, we see great promise in extending these methods to the problem of re-engagement: guiding interventions that increase the likelihood that volunteers return to the system after leaving.

## 8 Acknowledgments

This work was supported in part by EU FP7 FET SmartSociety project under Grant agreement no. 600854 and SOCIAM funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/J017728/1. We thank Mark Hartswood for valuable discussions, and Eran Yogev and Dor Amir for help with programming.

## References

- [Anderson *et al.*, 2013] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Steering user behavior with badges. In *Proceedings of the 22nd international conference on World Wide Web*, pages 95–106, 2013.
- [Anderson *et al.*, 2014] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Engaging with massive online courses. In *Proceedings of the 23rd international conference on World wide Web*, pages 687–698. ACM, 2014.
- [Eveleigh *et al.*, 2014] Alexandra Eveleigh, Charlene Jennett, Ann Blandford, Philip Brohan, and Anna L Cox. Designing for dabblers and deterring drop-outs in citizen science. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 2985–2994. ACM, 2014.
- [Horton and Chilton, 2010] John Joseph Horton and Lydia B Chilton. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM conference on Electronic commerce*, pages 209–218. ACM, 2010.
- [Horvitz and Apacible, 2003] Eric Horvitz and Johnson Apacible. Learning and reasoning about interruption. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 20–27. ACM, 2003.
- [Ipeirotis *et al.*, 2010] Panagiotis G Ipeirotis, Foster Provost, and Jing Wang. Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD workshop on human computation*, pages 64–67. ACM, 2010.
- [Ipeirotis, 2010] Panagiotis G Ipeirotis. Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students*, 17(2):16–21, 2010.
- [Kamar *et al.*, 2013] Ece Kamar, Yaakov Kobi Gal, and Barbara J Grosz. Modeling information exchange opportunities for effective human–computer teamwork. *Artificial Intelligence*, 195:528–550, 2013.
- [Ling *et al.*, 2005] Kimberly Ling, Gerard Beenen, Pamela Ludford, Xiaoqing Wang, Klarissa Chang, Xin Li, Dan Cosley, Dan Frankowski, Loren Terveen, Al Mamunur Rashid, et al. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10(4):00–00, 2005.
- [Mao *et al.*, 2013] Andrew Mao, Ece Kamar, and Eric Horvitz. Why stop now? predicting worker engagement in online crowdsourcing. In *First AAAI Conference on Human Computation and Crowdsourcing*, 2013.
- [Preece and Shneiderman, 2009] Jennifer Preece and Ben Shneiderman. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1):13–32, 2009.
- [Raddick *et al.*, 2010] M Jordan Raddick, Georgia Bracey, Pamela L Gay, Chris J Lintott, Phil Murray, Kevin Schawinski, Alexander S Szalay, and Jan Vandenberg. Galaxy zoo: Exploring the motivations of citizen science volunteers. *Astronomy Education Review*, 9(1):010103, 2010.
- [Reed *et al.*, 2013] Jeff Reed, M Jordan Raddick, Andrea Lardner, and Karen Carney. An exploratory factor analysis of motivations for participating in zooniverse, a collection of virtual citizen science projects. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, pages 610–619. IEEE, 2013.
- [Rogstadius *et al.*, 2011] Jakob Rogstadius, Vassilis Kostakos, Aniket Kittur, Boris Smus, Jim Laredo, and Maja Vukovic. An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. In *Intl. Conf. on Weblogs and Social Media*, 2011.
- [Sauermaann and Franzoni, 2015] Henry Sauermaann and Chiara Franzoni. Crowd science user contribution patterns and their implications. *Proceedings of the National Academy of Sciences*, 112(3):679–684, 2015.
- [Savage *et al.*, 2015] Saiph Savage, Andres Monroy-Hernandez, and Tobias Hollerer. Botivist: Calling volunteers to action using online bots. *arXiv preprint arXiv:1509.06026*, 2015.
- [Segal *et al.*, 2015] Avi Segal, Ya’akov Kobi Gal, Robert J Simpson, Victoria Victoria Homsy, Mark Hartswood, Kevin R Page, and Marina Jirotko. Improving productivity in citizen science through controlled intervention. In *Proceedings of the 24th International Conference on World Wide Web*, pages 331–337, 2015.
- [Shrot *et al.*, 2014] Tammar Shrot, Avi Rosenfeld, Jennifer Golbeck, and Sarit Kraus. Crisp: an interruption management algorithm based on collaborative filtering. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 3035–3044. ACM, 2014.
- [Simpson *et al.*, 2014] Robert Simpson, Kevin R Page, and David De Roure. Zooniverse: observing the world’s largest citizen science platform. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 1049–1054, 2014.
- [Tran-Thanh *et al.*, 2015] Long Tran-Thanh, Trung Dong Huynh, Avi Rosenfeld, Sarvapali D. Ramchurn, and Nicholas R. Jennings. Crowdsourcing complex workflows under budget constraints. In *AAAI*, 2015.