

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

MOBILIZING THE CITIZEN CROWD

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

by

Rajan Vaish

December 2015

The Dissertation of Rajan Vaish
is approved:

Professor James E. Davis, Chair

Professor Michael S. Bernstein

Professor Arnav Jhala

Dean Tyrus Miller
Vice Provost and Dean of Graduate Studies

Copyright © by

Rajan Vaish

2015

Table of Contents

List of Figures	vi
List of Tables	ix
Abstract	x
Dedication	xii
Acknowledgments	xiii
1 Introduction	1
1.1 Background and Limitations	4
1.1.1 Micro contributions: few seconds of light weight tasks	4
1.1.2 Large scale distributed and time-critical campaigns: few hours of complex tasks	6
1.1.3 Highly intellectual contributions at scale: few months of extremely expert tasks	7
1.2 Contribution	8
1.2.1 Twitch Crowdsourcing: Crowd Contributions in Short Bursts of Time	9
1.2.2 The Whodunit Challenge: Mobilizing the Crowd in India	10
1.2.3 The Aspiring Researchers Challenge: Crowdsourcing Research at Scale	11
2 Twitch Crowdsourcing: Crowd Contributions in Short Bursts of Time	13
2.1 Introduction	13
2.2 Related Work	17
2.2.1 Twitch: Census	17
2.2.2 Twitch: Photo Ranking	18
2.2.3 Twitch: Structuring the Web	18
2.3 Twitch Crowdsourcing	18
2.4 Twitch	21

2.4.1	Census	22
2.4.2	Photo Ranking	24
2.4.3	Structuring the Web	25
2.5	Field Deployment and Evaluation	28
2.5.1	Android Deployment	28
2.5.2	Field Experiments: Speed and Cognitive Load	32
2.6	Discussion	39
2.7	Conclusion	41
3	The Whodunit Challenge: Mobilizing the Crowd in India	43
3.1	Introduction	43
3.2	Related Work	46
3.3	The Whodunit Challenge	48
3.3.1	Design Principles	48
3.3.2	Contest Mechanics	49
3.3.3	Chance of Winning	51
3.3.4	Publicity and Outreach	52
3.4	Analysis Methodology	53
3.5	Results	54
3.5.1	Winning Strategies	55
3.5.2	Emergent Themes	60
3.6	Discussion and Recommendations	65
3.7	Conclusion	68
4	The Aspiring Researchers Challenge: Crowdsourcing Research at Scale	70
4.1	Introduction	70
4.2	Related Work	72
4.3	The Aspiring Researchers Challenge	73
4.3.1	Coordination Technique	74
4.3.2	The Process	76
4.3.3	Projects and Demographics	78
4.4	Research Components and Findings	80
4.4.1	Participation and Longevity	81
4.4.2	Paper Writing at Scale	91
4.4.3	Credit Distribution and Team Dynamics	95
4.5	Survey Findings and Discussion	101
4.6	Conclusion and Future Work	104
A	The Aspiring Researchers Challenge: Previous Work	106
A.1	Pilot Study 1	107
A.2	Pilot Study 2	108

B The Aspiring Researchers Challenge: Crowd Publications	110
B.1 Daemo: a Self-Governed Crowdsourcing Marketplace	110
B.1.1 Meta Information	110
B.1.2 Overview	111
B.2 On Optimizing Human-Machine Task Assignments	111
B.2.1 Meta Information	111
B.2.2 Overview	112
B.3 Investigating the "Wisdom of Crowds" at Scale	112
B.3.1 Meta Information	112
B.3.2 Overview	113
B.4 RTI Compression for Mobile Devices	114
B.4.1 Meta Information	114
B.4.2 Overview	114
C The Aspiring Researchers Challenge: HCI Project - Idea Flow	115
Bibliography	118

List of Figures

2.1	Twitch crowdsourcing asks for one to two seconds of users time each time they unlock their phone. From left to right: Census (crowd size), Census (attire), Photo Ranking, and Structuring the Web.	14
2.2	Subsamples from the worldwide Census activity generated by Twitch. Upper left: Cambridge, Massachusetts. Upper right: Bay Area, California. Bottom: worldwide.	19
2.3	Census tasks, activity and energy level.	21
2.4	Instant feedback shows people agreeing with the user in her location.	22
2.5	Verifying a web extraction animates a new branch onto the knowledge graph (left). Rejecting the extraction shows how far the user is in the article (right).	26
2.6	Response times decline over the first forty unlocks. The line indicates the median; band is the interquartile range.	34
3.1	Graphical illustration of the Whodunit Challenge.	48
3.2	Chance of finding N or more phrases as a function of team size.	51
3.3	Number of unique missed calls vs. time.	55
3.4	Heat map of received missed calls.	56
3.5	Top 20 teams in the Whodunit Challenge. (*) We asked teams to report the total number of phrases that they submitted without help from. (‡) Data not available. other teams.	58
4.1	Left - Screengrab of various video meeting sessions through Google Hangout on Air; Right - Screengrab of discussions using Slack - a common communication medium.	76
4.2	Crowd contributed to all the projects in a variety of ways - storytelling, prototyping, data analysis, brainstorming, peer feedback and voting to name a few. Above are a few screengrabs from some of these efforts by crowd.	78

4.3	Top left - screen grab of wisdomofcrowds.stanford.edu website, that was built as part of the Data Science project. Top right - results from the Computer Vision, showing a cost savings of 63% was achieved in the experiments. Bottom - screen grabs of the crowd platform built by crowd, shows the reputation system and task authoring interface from daemo.stanford.edu.	79
4.4	The measure of activity and participation in the HCI project: people reading and writing on Slack, collaborating towards discussions and paper writing. Slack was our primary channel of communication in the HCI project. Note the huge spike when HCI project recruited 373 more students in the second round.	83
4.5	The measure of activity and participation in the HCI project: all time messages from people on Slack. Note high activity around second round recruitment (12th week), UIST submission (19th week) and CHI submission (30th week).	84
4.6	The measure of activity and participation in the Computer Vision and Data Science projects: number of people reading on Piazza. Data Science project ran for 12 weeks, while Computer Vision lasted for 15 weeks.	85
4.7	The measure of activity and participation across all projects: meeting views on YouTube from one of week of hosting. Note a huge spike for the HCI project, caused by the second recruitment.	86
4.8	The measure of activity and participation across all projects: estimated minutes on YouTube from one of week of hosting. Note a huge spike for the HCI project, caused by the second recruitment. All the videos were unlisted, to prevent viewing by people outside our network. People who were unable to attend live meeting, viewed this video later.	87
4.9	The measure of activity and participation in the HCI project: number of Github commits. Note the spikes around 13th (second round of recruitment), 19th (around UIST deadline) and 26th week (preparing for studies for CHI deadline).	88
4.10	Top five reasons for people to drop-off in all three projects. Inability to follow the project after exams was the primary reason for Computer Vision and Data Science projects, while time commitment happens to be the primary reason for the HCI project. 'Other' reasons include: inability to understand the direction of the project, fatigue due to long duration, etc.	91
4.11	In the HCI project, crowd and faculty collaborated to write the CHI paper. This figure shows that crowd dominated the writing effort with about 4.5 times more edits than the advisor - a rare achievement, given the fact that more than 80% of the participants had never written a paper before.	92

4.12	In the HCI project, crowd and faculty collaborated to write the CHI paper. This figure shows the activity of crowd vs. advisor throughout the duration of paper writing process. Though crowd did most of the work, the advisor became extremely active around the final deadline - helping synthesize the paper.	93
4.13	In the HCI project, crowd and faculty collaborated to write the CHI paper. This figure shows section wise evolution of the paper - starting with 'overview' and ending with 'discussion' - a very standard approach to write a paper.	94
4.14	Credit distribution network graph in the HCI project for CHI paper: size of the nodes determine the PageRank, largest being the top author (node# = 38); width of the arrows determine the score sent from node A to B. Different color represent different groups, based on expertise/interest or cultural binding. Note higher scores exchanged between orange and red clusters among each other, while the largest node (node# = 38) receives scores of varying size from multiple sources.	97
4.15	Credit distribution network graph in the HCI project for UIST paper: size of the nodes determine the PageRank, largest being the top author (node# = 26); width of the arrows determine the score sent from node A to B. Different color represent different groups, based on expertise/interest or cultural binding. The largest node in red is an RA, who distributed badges to all and create an ecosystem for acknowledging via badges. Only one example of major score exchange is observed in the intra group in orange - significant reduction from CHI, possibly due to the public nature of distribution.	100
C.1	Flow of ideas in the HCI project, from 23 categories to 4, in 5 iterations (8 weeks). The project was to build a new crowd market, that gives more power and features to workers and requesters. Darker shades represent ideas that faculty echoed in weekly meetings, while lighter shades represent ideas that were not reflected or echoed. Number represents, number of times a particular idea was proposed. Some ideas lasted a few iterations and did not contribute to the final four (bottom three categories). Though "Micro and macro task" appears to be less popular (proposed fewer times), it consistently ranked higher (people voted it highly) - hence, it made it to the final four.	116

List of Tables

2.1	Field deployment Twitch unlock times (N=11,014). Answer times for Twitch usage in the wild.	33
2.2	3-back task performance (N=5405 tasks). Median and mean delay times for answering 3-back tasks in the cognitive load study. 2-back results are similar.	37
4.1	Percentage of people who completed or helped with the last milestones. Assuming that MOOCs take 12 weeks on average (it could be less), we can see that though our projects took longer for final completion (writing a paper vs. final exam) - the drop off rate was lower.	90

Abstract

Mobilizing the Citizen Crowd

by

Rajan Vaish

Crowdsourcing is a powerful approach to solve some of the biggest problems today - social and scientific, ranging from helping blind people [51] to advancing the state of computer vision [113]. To harness the potential of crowdsourcing, researchers have often relied and successfully used crowd marketplaces such as Amazon Mechanical Turk [1]. However, in spite of a large worker base and past successes, these platforms come with their own limitations. 40% of all the tasks are spam on Mechanical Turk [21]; and it's also difficult to accomplish tasks which require a broader participation base, or expert knowledge. Citizen crowd, at the same time, involves distributed group of people with diverse background and skills; who might organize to accomplish tasks or solve problems without engaging with paid crowd platforms. The motivation for them lies in reasons other than micro-payments.

In this thesis, I explore problems of varying size and duration, where mobilizing the citizen crowd can achieve worthwhile goals. To investigate the possibilities with citizen crowdsourcing, I built systems and ran initiatives that require different level of expertise. This thesis highlights three projects, ranging from seconds of micro contributions to months of highly intellectual contributions - exploring the limits beyond existing micro and macro work.

In the first chapter, I present a mobile lock screen application [126], that would help overcome the friction and channel factors for micro contributions. Anyone with the app can make light weight contributions every time they unlock their phones, in short bursts of time. In a span of six months, we registered more than 100,000 contributions.

In the second chapter, I discuss the Whodunit Challenge [127] - a large scale distributed and time-critical campaign we ran in India. As part of the initiative, we tried to understand crowd mobilization dynamics in the developing world. Our campaign required a few hours of complex coordination to execute the challenging goals; and in less than a day, we attracted about 10,000 participants who solved the task.

In the third chapter, I discuss the Aspiring Researchers Challenge [38, 125]- a first of its kind research-at-scale initiative, that provided access to more than 1,000 aspiring researchers from around the world to do academic style research. The initiative required months of highly intellectual contributions that usually experts like graduate students do, under engineering faculty at Stanford and UC Santa Cruz. Over the span of six months, these crowd researchers worked on three projects in computer science, and produced three work-in-progress papers at top-tier conferences [70, 117, 128]. One project's full paper is under submission, and two crowd participants earned a full time RA position at Stanford University. On the basis of this work, we conclude that it's possible to mobilize the citizen crowd to accomplish worthy goals.

To my parents and brother,
for their endless love, support and freedom to pursue my goals

To my wife,
for her care, encouragement and consistent motivation to dream big

Acknowledgments

First and foremost, I would love to thank my wonderful advisors James Davis and Michael Bernstein, who not only taught me how to do impactful research, but always supported me and my goals by all means. Their invaluable guidance and friendliness has made my graduate journey pleasant and rewarding. I can't thank them enough.

Special thanks to Arnav Jhala for serving on my dissertation committee; and Suresh Lodha and Sri Kurniawan for the advancement committee.

My graduate experience would be incomplete if I had not done four amazing internships. I would love to thank my incredible mentors Bill Thies and Ed Cutrell from Microsoft Research India; Michael Muller and Werner Geyer at IBM T.J. Watson Research Cambridge; Victoria Bellotti at PARC; and Andres Monroy-Hernandez and Jaime Teevan at Microsoft Research Redmond. At MSR Redmond, I was also fortunate to collaborate with other researchers who asked tough questions and helped me frame my research; special thanks to Shamsi Iqbal, Saleema Amershi, Ece Kamar, Dan Liebling and Kristina Toutanova. At these internships, I also had a very valuable experience of working with other PhD students - at MSR India, I worked with Aditya Vashistha from University of Washington; at PARC, I worked with Vera Liao from UIUC; at MSR Redmond, I worked with Carrie Cai from MIT, Niloufar Salehi from Stanford, Rhema Linder from Texas A&M and Ke Tran from University of Amsterdam - thanks all.

Over the past few years, I got a chance to collaborate with many people who helped me with my research. At UC Santa Cruz, I would like to thank Jing Liu, Steven

Scher, Prabath Gunawardane, Sascha T. Ishikawa, Bipeng Zhang, Jacob Thomas, Emily Lovell and Mrunal Gawade. At Stanford, I would like to thank Keith Wyngarden, Jingshu Chen and Brandon Cheung. At HCOMP CrowdCamp 2013, I had a great opportunity to collaborate with Jeffrey Bigham from Carneige Mellon, Haoqi Zhang from Northwestern, Peter Organisciak from UIUC and Kotaro Hara from UMD College Park - thanks to all.

As a graduate student, I co-organized a first of its kind “research at scale” initiative, where I collaborated with hundreds of students from around the world - huge thanks to all. This effort wouldn’t have been possible without the help of amazing collaborators - Sharad Goel, Camelia Simoiu, Imanol Ibarra, Geza Kovacs and Ranjay Krishna at Stanford University; and Serge Belongie, Andreas Veit and Michael Wilber at Cornell University/Cornell Tech.

My graduate journey won’t be complete without my friends by my side. I would like to thank Abhishek Tiwari, Manisha Rana, Miten Jain, Parameswaran Raman, Prakash Ganapathy, Prakhar Agarwal, Rahul Sharma, Rekha Pitchumani, Siddharth Dixit, Utkarsh Shrivastava and Yuvraj Negi the most.

Finally, I would like to thank my funding sources which made all of my research possible: the University of California Regents Fellowship and Reid Porter at Los Alamos National Lab through the ISSDM collaboration with UC Santa Cruz.

Chapter 1

Introduction

Crowdsourcing is an approach to harness the expertise or intelligence of a distributed group of people, to solve a problem or execute any given task. In recent years, the growth of internet and communication technologies have made collaboration easier and brought people together to achieve a variety of goals. These goals have ranged from improving search engine results [50], to digitizing books [131] to creating the world's largest encyclopedia [40] - varying across different duration, size and approach. In crowdsourcing, these approaches have been broadly defined into two types [91] - directive or marketplace crowdsourcing and collaborative or citizen crowdsourcing.

In marketplace crowdsourcing, researchers have successfully used the model of microtasking - breaking down huge tasks into smaller ones, and distributing them among the people all over the web via crowd marketplaces. These people are often termed as "workers", who make a variety of contributions for a small payment as instructed by "requesters", who design and post the tasks. This approach has helped solve

large scale problems like annotating large sets of images [61]; providing intelligence inside of interactive systems [51, 49], and fostered social science, artistic and educational research [91]. Among many crowd marketplaces, that lets requester architect and post such tasks, Amazons Mechanical Turk leads the domain with over 500K workers in 190 countries [2] . This provides great opportunities in the space of crowd work. However, in spite of a large worker base and past successes on existing microtask platforms, these platforms come with their own limitations. 40% of all the tasks are spam on Mechanical Turk [21]. It is also difficult to find groups of people with topical or local expertise; and therefore it is non trivial to accomplish tasks which require a broader participation base, or expert knowledge. Centralized structure, lack of regulation and trust also questions labor practices and work ethics [115].

In citizen crowdsourcing, the crowd involves distributed people, who might organize to accomplish tasks or solve problems without engaging with paid microtask platforms, and be motivated for reasons other than micro-payments. Instead, they are primarily motivated by incentives such as interest, global or personal good, career building, pleasure, social, rewards, or visibility. The range and flexibility which citizen crowds have to offer overcomes some of the limitations of microtask platforms, and also allows for expert engagement. Some of the most successful crowdsourced projects in past have involved mobilizing the citizen crowd, and have varied in size, level of crowd participation and duration - such as: DARPA Network Challenge [7], the Polymath project [59], Wikipedia [40], txtEagle [65], FoldIt [58] and Duolingo [130].

To understand the potential of the citizen crowd further, I have been exploring

problems, motivations and avenues where mobilizing the citizen crowd is important to achieve varied goals. In this thesis, I'll share my findings in the space of citizen crowdsourcing, specifically looking into crowd participation at different scales of duration and complexity - ranging from a few seconds of microtask to a few months of highly intellectual contribution. This dissertation essentially attempts to explore the limits of citizen crowdsourcing at both ends - micro and macro contribution for worthwhile goals, with minimal to highly intellectual contributions. This thesis introduces three projects that are initiatives and systems built to coordinate and mobilize crowd for different goals and scales.

First, I propose an approach to engage people who are pressed for time, yet want to make effortless contribution in science, technology and understanding of our environment. To help them make such contributions, we built an unlock screen app for mobile devices [126]. This app minimizes usability friction and encourages user to execute worthwhile micro tasks, every time they unlock their phones - in short bursts of time. In the span of six months from deployment, the app registered more than 100,000 contributions.

Second, I present a distributed and time-critical campaign [122], that mobilized the crowd throughout a large country, to accomplish a complex goal within a few hours. The initiative [127] was set up to run with minimal internet availability and required people to collaborate and work with each other to be successful. In less than a day, the campaign involved about 8,000 people from around the country.

Third, I reflect upon the experiences of a first of its kind initiative, that brought

academic style research at scale and required highly intellectual crowd contribution. The initiative provided educational and research opportunity to more than 1,000 people from 6 continents, where about 25% members were women. The initiative is still ongoing and is in its 6th month, helping accomplish large scale open-ended projects in the area of human-computer interaction, computer vision and data science. Over the span of six months, all three projects produced a work-in-progress paper at top tier computer science conferences [70, 117, 128].

1.1 Background and Limitations

The space of citizen crowdsourcing is a broad one. Over the past few years, several initiatives have attempted to solve a variety of problems in science, arts, education and research. These approaches vary in scale, time commitment and intellectual contributions. In this section, I'll provide background to three type of citizen crowdsourcing approaches that helps mobilizing the crowd and inspired my research.

1.1.1 Micro contributions: few seconds of light weight tasks

1.1.1.1 Overview

Micro contribution [56] is an extremely common approach in marketplace crowdsourcing, where crowd executes small tasks for a given payment. However, most of the tasks on these platforms are redundant and does not match workers interest or topic they would care about - often reflecting in bad quality work. To reach a larger

pool of people, who would care about a topic and produce high quality output, we rely on citizen crowdsourcing efforts like citizen science projects [119], ESP games [87] and games for purpose [129].

Through micro contributions, citizen science projects have fostered exploration and discovery of scientific importance such as: craters in mars [121], galaxies [107], protein structures [58] and RNA sequencing [86]. The impact goes beyond science, where citizens have rallied to take collective action and volunteered to participate in search mission of Malaysian Airlines [5], UC Berkeley researcher Jim Gray [75], and provided eyewitness reports during crises and violent government crackdowns [98].

1.1.1.2 Limitations

Though these approaches and projects have produced great results, they suffer from one limitation - channel factors [cite] or friction of usage. To make any contribution at the moment, a person has to go through a series of steps, like finding then launching an app, signing in, finding the task, etc. These seemingly small requirements are not naturalistically placed in a persons daily habit and can often deter a busy person to not contribute at all. Providing a means to overcome this limitation will be highly impactful for citizen crowdsourcing projects.

1.1.2 Large scale distributed and time-critical campaigns: few hours of complex tasks

1.1.2.1 Overview

In 2009, MIT created history, when it successfully solved the DARPA Network Challenge [7] within nine hours of launch, and utilized citizen crowdsourcing through internet and social media [122]. This was first of its kind time-critical and highly distributed initiative; that encouraged mobilization of crowd from around the United States to find ten red balloons, spread across the country. Its success proved that it is possible to accomplish possibly complex tasks like natural disaster response to quickly locating missing children within few hours; when strategized, collaborated, incentivised and communicated effectively.

This challenge opened avenues to solve a variety of problems, and more efforts were launched in the US and Europe. In 2011, University of Pennsylvania launched the MyHeartMap Challenge [26], where crowd was mobilized to find the location of life saving AEDs across Philadelphia. In 2012, the US Department of State launched the Tag Challenge [36], where crowd had to locate people of interest across five cities in United States, England, Sweden and Slovakia within a week.

1.1.2.2 Limitations

Though these efforts gave useful insights, they were still limited to situation and circumstances in the developed world. For a global impact, it was necessary to

understand the dynamics of the developing world, where internet availability is minimal and cultures vary. Exploring the possibilities and requirements of mobilizing crowd in a developing country will help in organizing such efforts on a global scale, and solve extremely large scale distributed problems.

1.1.3 Highly intellectual contributions at scale: few months of extremely expert tasks

1.1.3.1 Overview

There are several tasks, that last long and until recently were only possible to be done by experts - like authoring encyclopedia, writing stories and playing international level chess. Now, it is possible for anyone to participate in these initiatives [40, 81, 94] - thanks to growth of collaboration technologies and research in citizen crowdsourcing. However, one area that has been left unexplored is scientific collaboration. Scientific research is an extremely expert task, that is confined to researchers at universities and labs. The traditional approach to scientific research is robust and proven, however, it limits the scale of the project and access to people interested in pursuing it. Traditional practices also experience resource constraints such as - funding and diverse skill, that could be addressed with crowd participation.

In the spirit of collaboration, scientists around the world have begun sharing resources and their findings to further the science - projects like Open Science [104], MIT Registry of Standard Biological Parts [31] and UCSC Genomics [79] are a few examples in this space. Scientists and researchers are also beginning to write large

co-authored papers together, moving beyond Physics [27] to human-computer interaction [124]. However, these initiatives are among committed experts, where the cost of coordination and risk of execution is low. In past, projects like FoldIt [58] and EteRNA [86] have involved people from around the world to solve scientific problems through games. However, their contribution was not in intellectual capacity. Among all, there has been possibly one project that stands out - the Polymath project [59]. The project invited people to solve a specific complex mathematical problem, and though most of the participants were experts, it gave an opportunity and access to all.

1.1.3.2 Limitations

Though these efforts are promising and open new avenues, theres no initiative that provides access to non experts to make highly intellectual contributions for open ended problems in computer science. Working on open ended problems with people around the world is extremely non trivial and could have high coordination costs, but this space is left unexplored. Achieving this goal and understanding its approach can open new avenues for discoveries and produce upcoming researchers.

1.2 Contribution

The contribution of my dissertation is to address the limitations of past approaches in citizen crowdsourcing that vary across complexity and duration. Through my work, I investigate and attempt to stretch the boundaries of citizen crowdsourcing - by making micro work even smaller and frictionless; and making macro work

even larger with highly intellectual contribution requirement. My research explores opportunities from few seconds of light weight tasks to few months of extremely expert tasks. Through three projects I worked on, I attempt to: 1) minimize friction for micro contributions through Twitch crowdsourcing unlock screen app, 2) develop an understanding of crowd mobilization during time-critical campaigns in the developing world - by launching Whodunit Challenge in India, 3) solve large scale open-ended research problems by providing access to people around the world to work with experts - by launching the Aspiring Researchers Challenge.

1.2.1 Twitch Crowdsourcing: Crowd Contributions in Short Bursts of Time

To lower the threshold to participation in crowdsourcing, we present twitch crowdsourcing: crowdsourcing via quick contributions that can be completed in one or two seconds [126]. We introduce Twitch, a mobile phone application that asks users to make a micro-contribution each time they unlock their phone. Twitch takes advantage of the common habit of turning to the mobile phone in spare moments. Twitch crowdsourcing activities span goals such as authoring a census of local human activity, rating stock photos, and extracting structured data from Wikipedia pages. At the time of CHI'14 paper submission, we report a field deployment of Twitch - where 82 users made 11,240 crowdsourcing contributions as they used their phone in the course of everyday life. After six months of deployment, over 100,000 contributions were registered. The median Twitch activity took just 1.6 seconds, incurring no statistically distinguish-

able costs to unlock speed or cognitive load compared to a standard slide-to-unlock interface.

1.2.2 The Whodunit Challenge: Mobilizing the Crowd in India

While there has been a surge of interest in mobilizing the crowd to solve large-scale time-critical challenges, to date such work has focused on high income countries and Internet-based solutions. In developing countries, approaches for crowd mobilization are often broader and more diverse, utilizing not only the Internet but also face-to-face and mobile communications. In this paper [127], we describe the Whodunit Challenge, the first social mobilization contest to be launched in India. The contest enabled participation via basic mobile phones and required rapid formation of large teams in order to solve a fictional mystery case. The challenge encompassed 7,700 participants in a single day and was won by a university team in about 5 hours. To understand teams strategies and experiences, we conducted 84 phone interviews. While the Internet was an important tool for most teams, in contrast to prior challenges we also found heavy reliance on personal networks and offline communication channels. We synthesize these findings and offer recommendations for future crowd mobilization challenges targeting low-income environments in developing countries.

1.2.3 The Aspiring Researchers Challenge: Crowdsourcing Research at Scale

Research is a high skill and resource intensive activity, both in time and effort, and often follows an ad hoc process. In a research process, it's often unclear what process, which if repeated produces a publishable paper. In the research ecosystem, experienced researchers with novel ideas are constrained with limited resources; and motivated people with exceptional skill lack direction or access to a research mentor.

In this project, I introduce "The Aspiring Researchers Challenge" [38, 125], an experiment in massive open online research, that explores the possibility of research at scale by connecting an expert with crowd (aspiring researchers). In span of six months, we had more than 1,000 sign ups from people around the world with almost no research experience - more than 80% had never published a paper before, about 25% were female and more than 70% were undergraduates.

We developed a weekly structure that helped in coordinating crowd, distribute fair credits, educate and train them about research topics and process. The structure utilized peer review to scale the process, that would include series of research phases like, brainstorming, prototyping, development and user-evaluation. Through our approach, participants worked on projects in computer vision, data science and human-computer interaction (HCI) - mentored by professors at Stanford University and UC Santa Cruz. And were able to successfully publish three work-in-progress papers at top-tier conferences in computer science - ACM UIST and AAAI HCOMP [70, 117, 128]. In the HCI

project, a full paper is under review at ACM CHI; while two participants who aspire for graduate school, earned a full time RA position at Stanford University. We believe that the project is a first step towards realizing research on a wider scale.

Chapter 2

Twitch Crowdsourcing: Crowd Contributions in Short Bursts of Time

2.1 Introduction

Mobilizing participation is a central challenge for every crowdsourcing campaign. Campaigns that cannot motivate enough participants will fail [105]. Unfortunately, many interested contributors simply cannot find enough time: lack of time is the top reason that subject experts do not contribute to Wikipedia [123]. Those who do participate in crowdsourcing campaigns often drop out when life becomes busy [90]. Even seemingly small time requirements can dissuade users: psychologists define channel factors as the small but critical barriers to action that have a disproportionate effect on whether people complete a goal [112]. Despite this constraint, many crowdsourcing campaigns assume that participants will work for minutes or hours at once, leading to

a granularity problem [46] where task size is poorly matched to contributors opportunities. We speculate that a great number of crowdsourcing campaigns will struggle to succeed as long as potential contributors are deterred by the time commitment.



Figure 2.1: Twitch crowdsourcing asks for one to two seconds of users time each time they unlock their phone. From left to right: Census (crowd size), Census (attire), Photo Ranking, and Structuring the Web.

To engage a wider set of crowdsourcing contributors, we introduce twitch crowdsourcing: interfaces that encourage contributions of a few seconds at a time. Taking advantage of the common habit of turning to mobile phones in spare moments [100], we replace the mobile phone unlock screen with a brief crowdsourcing task, allowing each user to make small, compounded volunteer contributions over time. In contrast, existing mobile crowdsourcing platforms (e.g., [65, 72, 96]) tend to assume long, focused runs of work. Our design challenge is thus to create crowdsourcing tasks that operate in very short time periods and at low cognitive load.

To demonstrate the opportunities of twitch crowdsourcing, we present Twitch,

a crowdsourcing platform for Android devices that augments the unlock screen with 1-3 second volunteer crowdsourcing tasks (Figure 2.1). Rather than a typical slide-to-unlock mechanism, the user unlocks their phone by completing a brief crowdsourcing task. Twitch is publicly deployed and has collected over eleven thousand volunteer contributions to date. The system sits aside any existing security passcodes on the phone.

Twitch crowdsourcing allows designers to tap into local and topical expertise from mobile users. Twitch supports three unlock applications:

1) *Census* envisions a realtime people-centered world census: where people are, what they are doing, and how they are doing it. For example, how busy is the corner caf at 2pm on Fridays? Census answers these questions by asking users to share information about their surroundings as they navigate the physical world, for example the size of the crowd or current activities (Figure 2.1).

2) *Photo Ranking* captures users opinion between two photographs. In formative work with product designers, we found that they require stock photos for mockups, but stock photo sites have sparse ratings. Likewise, computer vision needs more data to identify high-quality images from the web. Photo Ranking (Figure 2.1) asks users to swipe to choose the better of two stock photos on a theme, or contribute their own through their cell phone camera.

3) *Structuring the Web* helps transform the written web into a format that computers can understand. Users specify an area of expertise HCI, the Doctor Who television series, or anything else of interest on Wikipedia and help verify web ex-

tractions relevant to that topic. Each unlock involves confirming or rejecting a short extraction. In doing so, users could power a fact-oriented search engine that would directly answer queries like heuristic evaluation creator.

After making a selection, Twitch users can see whether their peers agreed with their selection. In addition, they can see how their contribution is contributing to the larger whole, for example aggregate responses on a map (Figure 2.2) or in a fact database (Figure 2.5).

We deployed Twitch publicly on the web and attracted 82 users to install Twitch on their primary phones. Over three weeks, the average user unlocked their phone using Twitch 19 times per day. Users contributed over 11,000 items to our crowdsourced database, covering several cities with local census information. The median Census task unlock took 1.6 seconds, compared to 1.4 seconds for a standard slide-to-unlock gesture. Secondary task studies demonstrated that Twitch unlocks added minimal cognitive load to the user.

Our work indicates that it may be possible to engage a broad set of new participants in crowdsourcing campaigns as they go about their day or have a few spare moments. In the following sections, we introduce twitch crowdsourcing in more detail and report on our public deployment and field experiments.

2.2 Related Work

Crowdsourcing has demonstrated the potential to support goals ranging from search engine support [50], classifier training [120], and even real-time interaction [48, 85]. However, these systems often are limited in the amount of topical and local expertise they can assume. To address this limitation, txtEagle [65], mCrowd [134], and mClerk [72] have sought to engage new sets of users through mobile crowdsourcing. However, these mobile crowdsourcing applications assume longer stretches of participation from crowd workers; twitch crowdsourcing aims to capture more incidental, ad-hoc opportunities. Bentley et al. characterize micro-moments by short (\sim 10 second) time periods when attention can be taken away from a current task [43, 47]. Twitch crowdsourcing demonstrates that these micro-moments are ideal platforms for interaction. Applications such as adaptive flashcards [66] operate on similar insights, but Twitch is unique in its design as an unlock screen.

2.2.1 Twitch: Census

Previous efforts have spread sensor-oriented crowdsourcing tasks through mobile phones [134]. Crowd-sensing applications likewise draw on humans-in-the-loop to trigger sensing actions or review results [106]. Scaling up sensor networks, MetroSense provides a people-centric paradigm for urban sensing [53]. Census also complements sensor technologies to track transportation habits [69] and mobile phone-based remote sensing platforms [60].

2.2.2 Twitch: Photo Ranking

User-generated photo voting platforms often succeed through a motivating goal, such as Kitten War [20]. However, while stock photography portals such as iStockPhoto [19] exist, none provide a vast collection of free to download images. In parallel, computer vision has demonstrated the importance of crowdsourcing large datasets [120]. Rankr [88] aims to rank images, ideas and priorities via a mobile phone application; however, it requires dedicated users and is not integrated into users general flow. Photo Ranking is also inspired by the Matchin game, which demonstrates that twitch crowdsourcing can complement game strategies [73].

2.2.3 Twitch: Structuring the Web

Campaigns such as FreeBase [9] and dbpedia [8] have demonstrated the opportunities available when crowds help structure the information available on the web. Automatic information extraction techniques such as ReVerb [67] can be augmented by crowd verification [77]. Twitch generalizes these interactive techniques and allows users to contribute to essentially any topic of interest.

2.3 Twitch Crowdsourcing

We address difficulties in mobilizing crowd participation by tapping into brief moments of users time. While waiting for an elevator, a meeting, a walk signal, or cafeteria service, it can be difficult to resist glancing at the phone [100]. Rather than



Figure 2.2: Subsamples from the worldwide Census activity generated by Twitch. Upper left: Cambridge, Massachusetts. Upper right: Bay Area, California. Bottom: worldwide.

making this temptation harder or less desirable, we take advantage of these short bursts to encourage people to make small contributions towards solving impactful problems. We call this approach twitch crowdsourcing: engaging with participants in a convenient fashion in short bursts.

While there are many possible short bursts of time in users daily lives, we focus on one that users already expect to be a small task: the mobile phone unlock screen. A slide-to-unlock gesture is useful for preventing accidental activation of the phone, but there is no reason why the unlock needs to be a slide gesture. We replace the unlock screen with a small crowdsourced task that takes roughly the same amount of time to complete. In doing so, the user contributes to crowdsourcing goals they want to support without taking extra time out of their day.

Twitch crowdsourcing interactions are very brief, so it is important that users can complete the tasks extremely quickly. Likewise, crowdsourcing is not the users primary task, so these tasks must be lightweight and not distracting. We frame the design of our twitch crowdsourcing applications around the following principles:

- Availability: summon the task with a single button press
- Quick completion: contribute within a couple of seconds
- Low cognitive load: do not distract users from their primary task
- Value: volunteer toward a worthwhile goal
- Personal impact: maximize the uniqueness of the users contributions [45]
- Longevity: keep tasks interesting after many repetitions
- Feedback: situate the users effort as directly contributing to the larger crowdsourcing goal
- Privacy: do not circumvent standard phone security

To maximize value and personal impact, we focus not on paid crowdsourcing (e.g., Amazon Mechanical Turk) but on activities to which the user might want to volunteer their time. Because twitch crowdsourcing happens while users are mobile, tasks that take advantage of the users local, physical context are particularly attractive. In addition, by lowering the threshold to contribution, we hypothesize that even busy topical experts might be able to contribute. Thus, we also explore opportunities to capture expertise, much like Wikipedia.

2.4 Twitch

Twitch is an Android application that appears when the user presses the phones power/lock button (Figures 2.1 and 2.3). When the user completes the twitch crowdsourcing task, the phone unlocks normally. Each task involves a choice between two to six options through a single motion such as a tap or swipe.

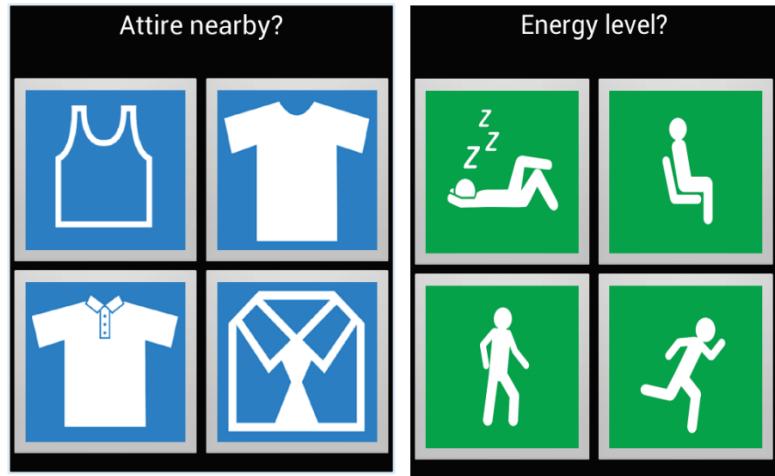


Figure 2.3: Census tasks, activity and energy level.

To motivate continued participation, Twitch provides both instant and aggregated feedback to the user. An instant feedback display shows how many other users agreed via a fadeout as the lock screen disappears (Figure 2.4) or how the users contributions apply to the whole (Figure 2.5). Aggregated data is also available via a web application, allowing the user to explore all data that the system has collected. For example, Figure 2.2 shows a human generated map from the Census application. To address security concerns, users are allowed to either disable or keep their existing An-

droid passcode while using Twitch. If users do not wish to answer a question, they may skip Twitch by selecting Exit via the options menu. This design decision has been made to encourage the user to give Twitch an answer, which is usually faster than exiting. Future designs could make it easier to skip a task, for example through a swipe-up.

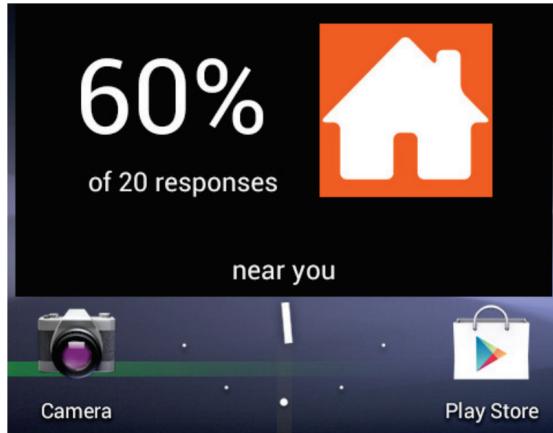


Figure 2.4: Instant feedback shows people agreeing with the user in her location.

Below, we introduce the three main crowdsourcing applications that Twitch supports. The first, Census, attempts to capture local knowledge. The following two, Image Voting and Structuring the Web, draw on creative and topical expertise. These three applications are bundled into one Android package, and each can be accessed interchangeably through Twitch's settings menu.

2.4.1 Census

Despite progress in producing effective understanding of static elements of our physical world routes, businesses and points of interest we lack an understanding of

human activity. How busy is the corner cafe at 2pm on Fridays? What time of day do businesspeople clear out of the downtown district and get replaced by socializers? Which neighborhoods keep high-energy activities going until 11pm, and which ones become sleepy by 6pm? Users could take advantage of this information to plan their commutes, their social lives and their work.

Existing crowdsourced techniques such as Foursquare are too sparse to answer these kinds of questions: the answers require at-the-moment, distributed human knowledge. We envision that twitch crowdsourcing can help create a human-centered equivalent of Google Street View, where a user could browse typical crowd activity in an area. To do so, we ask users to answer one of several questions about the world around them each time they unlock their phone. Users can then browse the map they are helping create.

Census is the default crowdsourcing task in Twitch. It collects structured information about what people experience around them. Each Census unlock screen consists of four to six tiles (Figures 2.1 and 2.3), each task centered around questions such as:

- How many people are around you?
- What kinds of attire are nearby people wearing?
- What are you currently doing?
- How much energy do you have right now?

While not exhaustive, these questions cover several types of information that a local census might seek to provide. Two of the four questions ask users about the people around them, while the other two ask about users themselves; both of which they are uniquely equipped to answer. Each answer is represented graphically; for example, in case of activities, users have icons for working, at home, eating, travelling, socializing, or exercising.

To motivate continued engagement, Census provides two modes of feedback. Instant feedback (Figure 2.4) is a brief Android popup message that appears immediately after the user makes a selection. It reports the percentage of responses in the current time bin and location that agreed with the user, then fades out within two seconds. It is transparent to user input, so the user can begin interacting with the phone even while it is visible. Aggregated report allows Twitch users to see the cumulative effect of all users behavior. The data is bucketed and visualized on a map (Figure 2.2) on the Twitch homepage. Users can filter the data based on activity type or time of day.

2.4.2 Photo Ranking

Beyond harnessing local observations via Census, we wanted to demonstrate that twitch crowdsourcing could support traditional crowdsourcing tasks such as image ranking (e.g., Matchin [73]). Needfinding interviews and prototyping sessions with ten product design students at Stanford University indicated that product designers not only need photographs for their design mockups, but they also enjoy looking at the

photographs. Twitch harnesses this interest to help rank photos and encourage contribution of new photos. *Photo Ranking* crowdsources a ranking of stock photos for themes from a Creative Commons-licensed image library. The Twitch task displays two images related to a theme (e.g., Nature Panorama) per unlock and asks the user to slide to select the one they prefer (Figure 2.1). Pairwise ranking is considered faster and more accurate than rating [73]. The application regularly updates with new photos.

Users can optionally contribute new photos to the database by taking a photo instead of rating one. Contributed photos must be relevant to the days photo theme, such as Nature Panorama, Soccer, or Beautiful Trash. Contributing a photo takes longer than the average Twitch task, but provides an opportunity for motivated individuals to enter the competition and get their photos rated.

Like with Census, users receive instant feedback through a popup message to display how many other users agreed with their selection. We envision a web interface where all uploaded images can be browsed, downloaded and ranked. This data can also connect to computer vision research by providing high-quality images of object categories and scenes to create better classifiers.

2.4.3 Structuring the Web

Search engines no longer only return documents they now aim to return direct answers [50, 57]. However, despite massive undertakings such as the Google Knowledge Graph [13], Bing Satori [22] and Freebase [9], much of the knowledge on the web remains unstructured and unavailable for interactive applications. For example, searching for

“Weird Al Yankovic born” in a search engine such as Google returns a direct result “1959” drawn from the knowledge base; however, searching for the equally relevant “Weird Al Yankovic first song”, “Weird Al Yankovic band members”, or “Weird Al Yankovic bestselling album” returns a long string of documents but no direct answer, even though the answers are readily available on the performer’s Wikipedia page.

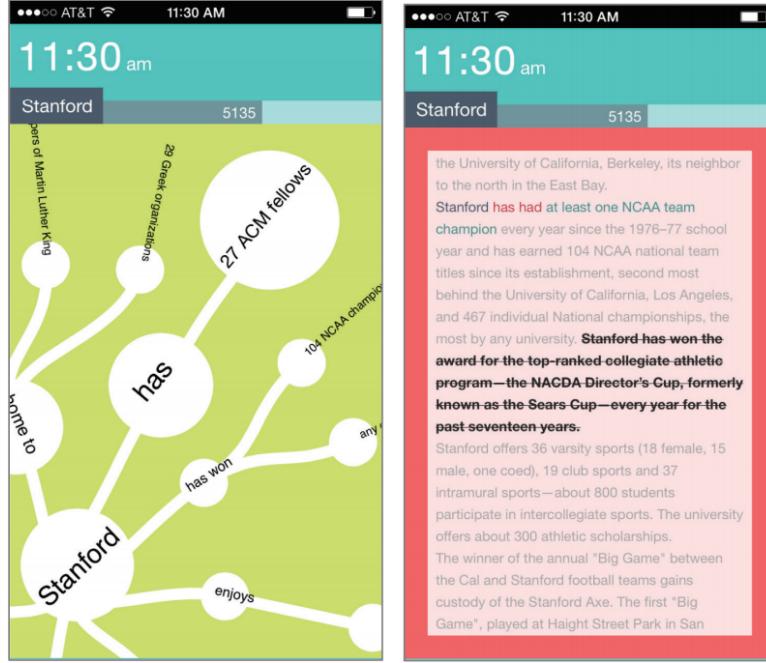


Figure 2.5: Verifying a web extraction animates a new branch onto the knowledge graph (left). Rejecting the extraction shows how far the user is in the article (right).

To enable direct answers, we need structured data that is computer-readable. While crowdsourced undertakings such as Freebase and dbpedia have captured much structured data, they tend to only acquire high-level information and do not have enough contributors to achieve significant depth on any single entity. Likewise, while information extraction systems such as ReVerb [67] automatically draw such information from

the text of the Wikipedia page, their error rates are currently too high to trust. Crowd-sourcing can help such systems identify errors to improve future accuracy [77]. Therefore, we apply twitch crowdsourcing to produce both structured data for interactive applications and training data for information extraction systems.

Contributors to online efforts are drawn to goals that allow them to exhibit their unique expertise [45]. Thus, we allow users to help create structured data for topics of interest. The user can specify any topic on Wikipedia that they are interested in or want to learn about, for example HCI, the Godfather films, or their local city. To do so within a one-to-two second time limit, we draw on mixed-initiative information extraction systems (e.g., [77]) and ask users to help vet automatic extractions.

When a user unlocks his or her phone, Structuring the Web displays a high-confidence extraction generated using ReVerb, and its source statement from the selected Wikipedia page (Figure 2.1). The user indicates with one swipe whether the extraction is correct with respect to the statement. ReVerb produces an extraction in Subject-Relationship-Object format: for example, if the source statement is Stanford University was founded in 1885 by Leland Stanford as a memorial to their son, ReVerb returns Stanford University, was founded in, 1885 and Twitch displays this structure. To minimize cognitive load and time requirements, the application filters only include short source sentences and uses color coding to match extractions with the source text.

In *Structuring the Web*, the instant feedback upon accepting an extraction shows the user their progress growing a knowledge tree of verified facts (Figure 2.5). Rejecting an extraction instead scrolls the user down the article as far as their most

recent extraction source, demonstrating the users progress in processing the article. In the future, we envision that search engines can utilize this data to answer a wider range of factual queries.

2.5 Field Deployment and Evaluation

This paper hypothesizes that short bursts of crowdsourcing can be embedded in users everyday activities. We focus our evaluation on the two main components of this claim. First, *is it possible to crowdsource meaningful tasks in short bursts of time?* Can Twitch attract sufficient volunteer usage to achieve its goals via short interaction bursts? Second, *it is possible to embed these tasks in users everyday activities?* Are the tasks sufficiently quick and easy that users can complete them without excessive burden?

To answer these questions, we pursued two evaluation strategies. First, we publically deployed Twitch and attracted 82 volunteer users over one month. We used this data to investigate naturalistic usage, collect crowdsourced data, and gather qualitative feedback. Second, we performed two field experiments on Twitch users to compare Twitchs time cost to a standard slide-to-unlock gesture, and a study of cognitive load using a working memory instrument called the 2- and 3-back test [99].

2.5.1 Android Deployment

To examine Twitchs ability to attract naturalistic crowdsourced data, we began with a public deployment. In doing so, we focused on unlock statistics, timing data,

location demographics, and duration of usage.

2.5.1.1 Method

We released Twitch on Google Play, a public marketplace for Android applications. We then used social networks and email lists to spread word of the application to individuals who would be willing to try the research prototype. This process resulted in 82 volunteer users installing the application and completing at least one Twitch task. While Twitch contained all of the applications, its focus and default setting was Census.

The application saved a timestamp each time a user pressed the power/unlock button and launched a Twitch unlock, and another timestamp each time the user completed a Twitch task. To address privacy concerns, we tied all data to a randomly generated phone identifier. Finally, to complement our quantitative metrics, we emailed all users of Twitch roughly three weeks after they began using the application with an invitation to complete a brief survey about their experiences.

2.5.1.2 Results

During the study period, 82 participants completed 11,240 Twitch tasks to unlock their phones. The task completion rate was 37.4%; other times, users pressed the power button but did not complete the task, most likely just checking the time or notifications without unlocking the phone.

Active Twitch users kept using the application for a significant length of time. As expected for free smartphone applications on a public marketplace, many window

shoppers dropped off early: about half (46.4%) of those who downloaded Twitch within the study period uninstalled it within a day. However, among the remaining users who kept the application for at least one day, the average duration of use was 31.9 days (median 30.9 days). One month of usage indicates a significant time and effort investment from these volunteers.

Census, the default Twitch application, received by far the most attention, with 9,717 responses from around the world (Figure 2.2), including the United States, Europe, India and Asia. The cities with the greatest number of responses were Cambridge, MA; Stanford, CA; and Pittsburgh, PA. Large numbers of responses also came from India and Japan. The resulting data is quite densely populated (Figure 2.2). The data indicates insights such as when cafes shift from work activities to social life, how many people are in a building early in the morning and late at night, and where people who work in downtown tend to go to exercise.

Roughly one fifth of users switched to try the Structuring the Web application, resulting in 334 responses. Across a range of topics such as “Google”, “Android”, “Earth”, and “United States”, Twitch users evaluated 181 unique statements from Wikipedia run through ReVerb. Users marked 41.4% of total extractions as correct.

At the end of the field study, we sent an online survey to 57 users who optionally provided their emails when installing the application. Sixteen participants filled out the survey in three days, with thirteen participants from the United States and three participants from Panama, Canada, and India. Participants generally enjoyed using Census (median Likert score 5 of 7), and found it only slightly distracting (median

= 3.5). Participants who self-selected to switch to Image Ranking enjoyed it as well (median = 6).

Twitch discourages skipping tasks - it involves a two-tap exit via a context menu - making it possible that some users would deliberately answer incorrectly just to quickly unlock their phone. However, most participants reported making their selections honestly and accurately (median = 6). Roughly half of participants (56%) never felt the need for a skip button or gesture.

In order to motivate our users to continue using Twitch, and to encourage longevity, we have complemented all the applications with immediate feedback that tells users the percentage of others agreeing. Users moderately enjoyed the popup feedback (median = 5.5). In addition, we provided a web interface where users are able to view their data on a crowdsourced activity map; participants reported that the site had some effect on motivating them to contribute further (median = 5).

Open ended responses reinforced that the application was “*very lightweight [] I’m sure I answered over 100 questions effortlessly.*” One area of improvement: when users were not moving, Census tasks in particular became repetitive and thus annoying. Some participants wished the immediate feedback to be richer or less precisely tuned to their location so that it reports more data. A few participants were conscious of their location identification. Those users who used Image Ranking enjoyed the experience, but as the application switching was not particularly discoverable (a design decision to make the public application simpler), many werent aware of its existence. One of the users shared her experience: “*The beautiful pictures brought happiness to the moment*”.

Most participants felt that concerns about privacy were addressed (median = 5); however, some users uninstalled the application due to privacy concerns with location tracking. Redundant tasks and privacy concerns were the core reasons reported for uninstalling the application.

Participants biggest requests for improving Twitch involved adding variety to the tasks or making tasks appear less frequently. Others wanted more intelligent automatic task selection: we believe this may be possible using the onboard phone sensors.

2.5.2 Field Experiments: Speed and Cognitive Load

For twitch crowdsourcing to succeed, tasks must be quick and low in cognitive load. If a task is too slow, users will opt out or give quick, unthoughtful responses. Likewise, if a task requires too much mental effort, users will have to disengage from whatever they are doing just to unlock their phones, which is undesirable. In a pair of field experiments tied to our deployment, we investigate the impact that Twitch has on the speed of unlocking users phones and on cognitive load.

2.5.2.1 Method: Speed Experiment

To compare unlock times to a standard unlock interface; we included a simple slide-to-unlock screen in Twitch when the application was publicly deployed. Slide-to-unlock is similar to the default Android screen lock (called Slider). Thus, slide-to-unlock acts as a speed baseline to compare with rest of the Twitch applications. Once

a week, Twitch displayed the slide-to-unlock screen instead of the selected Twitch task, along with a message stating that this temporary unlock interface was part of our field experiment. We compare user speed on the slide-to-unlock test against the rest of Twitch applications, examining whether Twitch tasks are noticeably slower or faster than a standard Android unlock. Speed is calculated in milliseconds from the moment Twitch is visible until the user completes the task.

2.5.2.2 Results

In naturalistic usage via a dataset of 11,014 unlocks (Table 2.1), slide-to-unlock took users a median of 1.39 seconds to unlock their phone (average 1.89). Median Census unlocks took between 1.491.89 seconds (average 2.012.49 seconds). Photo Ranking took 1.58 seconds (average 2.44). The median Structuring the Web unlock took 2.09 seconds (average 2.98 seconds), as it requires more reading.

Activity	Median	Mean
Slide-to-unlock	1421 ms	1956 ms
Census: people	1583 ms	2077 ms
Census: activity	1910 ms	2077 ms
Census: energy	1688 ms	2177 ms
Census: dress	1499 ms	2033 ms
Photo ranking	1579 ms	2438 ms
Structuring the web	2900 ms	3510 ms

Table 2.1: Field deployment Twitch unlock times (N=11,014). Answer times for Twitch usage in the wild.

We next tested whether the Twitch unlocks were in fact slower than a standard slide-to-unlock interface. As is common with reaction time data, Twitch duration data

are not normally distributed; even after transformation, the data were heteroskedastic and thus not amenable to parametric statistics. Thus, we use the Friedman test on medians, a nonparametric equivalent of the ANOVA with blocked designs. We compared median unlock times between Twitch tasks for users who completed at least 100 unlocks and tried all Census activities, excluding Structuring the Web and Photo Ranking because fewer users had tried it ($N=22$ users). The Friedman test revealed a significant effect of Twitch activity on unlock time ($\chi^2(4)=24.4$, $p <.001$). Post-hoc paired Wilcoxon signed rank tests using Bonferroni correction between conditions revealed that the “current activity” Census task was significantly slower than all other Census tasks and slide-to-unlock (all $p <.05$), but there were no other significant differences between conditions.

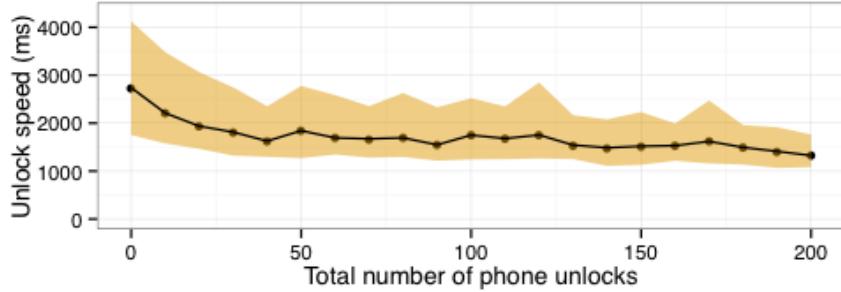


Figure 2.6: Response times decline over the first forty unlocks. The line indicates the median; band is the interquartile range.

Thus, while Twitch tasks may be slightly slower than a standard slide-to-unlock gesture (100-300ms), the difference is so slight in comparison to the variance that the two are not currently statistically distinguishable. There is clearly variability across Twitch tasks: Censuss current activity task is 300-400ms slower than the others,

possibly because it is the only categorical (non-scalar) response amongst the Twitch tasks, which impacts visual search.

Users quickly acclimated to Twitch, with median response time dropping from 2.7 seconds in the first ten unlocks to 1.6 seconds by the fortieth unlock (Figure 2.6). Given that users are unlocking their phones many times each day, they approach optimal performance quickly.

2.5.2.3 Method: Cognitive Load Experiment

To understand whether twitch might distract or annoy users, we turn to measures of cognitive load. In particular, we adopt a working memory task known as the 2-back and 3-back test [78, 99]. The 2-back test displays a stream of English letters, one at a time, and the participant must indicate for each letter whether it is the same as the letter that was displayed two letters ago. For example, if the previous letter stream were E, U, S, I, and the current letter were Y, the correct answer would be “Different”; if the current letter were S, the answer would be “Same”. Following previous work [78], we sample so that the correct answer is “Same” with probability 1/7. The 3-back test is identical to the 2-back test except the distance is three letters instead of two, making the task more difficult. To further load working memory, we likewise inject a Twitch Census activity or a Photo Ranking activity with probability 1/7 between letters. Twitch activities add to cognitive load but do not impact the correct answer they do not count as a letter.

We performed a within-subjects field experiment to test how much each Twitch

activity would negatively impact performance on the memory task. For each round of the study, the participant performed a series of 2- or 3-back tasks with random Twitch activity injections until they had completed all of the Twitch activities. One study factor was Twitch activity: we randomized the order of Twitch unlock screens that got injected into the task stream, including a slide-to-unlock control, for each round of the study. The second factor was memory task: participants rotated through rounds of 2-back or 3-back tasks until all Twitch activities had been tested in each condition. We randomized whether 2-back or 3-back tasks appeared first for each participant. The study continued for six rounds, three in each memory task condition. The study typically took fifteen minutes.

We recruited participants from the set of active Twitch users, so they would already be familiar with the applications. In exchange, participants received a raffle ticket for an Android tablet. Drawing on prior work [100, 135], we instructed participants to go on a walk outdoors following a predefined path at one of two major universities or (alternatively) following any safe route in their neighborhood. In forcing participants to walk, we made sure that users would need to look around to answer Census questions.

We measured reaction time and task accuracy on the memory task. In particular, we grouped each reaction time and accuracy measurement with the Twitch activity that occurred just before the task. Thus, we measured how much of an impact each unlock screen had on the next 2- or 3-back instance, compared to baselines of no unlock (3-back instance preceded by a 3-back instance) and slide-to-unlock.

Our hypothesis was that Twitch would incur a cost to working memory accu-

racy and reaction time compared to a slide-to-unlock or no unlock gesture. However, we hypothesized that this cost for most Twitch tasks would be similar to the cost of a simple slide-to-unlock.

2.5.2.4 Results

Fourteen Twitch users completed the study. The population was worldwide, centered in areas with the most Twitch users. Participants completed 5,405 back tasks.

As with many reaction time measurements, timing data was non-normal after dropping outliers. Unlike the previous study, it was possible to retain homogeneity of variance and approach normality by applying a transformation. Such transformations are common when analyzing reaction time data. We use a generalized power transformation called a Box-Cox transformation, which algorithmically searches for the power λ that best fits a normal distribution. For our data, $\lambda = -.25$, a reciprocal fourth root transformation of the data. We transform timing data for our statistical analysis, but report raw milliseconds to aid interoperability.

Activity	Median	Mean	Accuracy
Back task	1194 ms	1802 ms	90%
Slide-to-unlock	1724 ms	2228 ms	87%
Photo ranking	1506 ms	2009 ms	86%
Census: people	1737 ms	2332 ms	87%
Census: activity	1679 ms	2000 ms	93%
Census: energy	1881 ms	2238 ms	93%
Census: dress	2229 ms	2445 ms	95%

Table 2.2: 3-back task performance (N=5405 tasks). Median and mean delay times for answering 3-back tasks in the cognitive load study. 2-back results are similar.

The median task took 1.2 seconds to answer for both the 2-back and 3-back conditions when no Twitch activities were interposed (mean 1.68 for 2-back and 1.80 for 3-back, Table 2.2). Slide-to-unlock caused the task to take a median 1.7 seconds (mean 2.2). All Twitch tasks except Census clothing had similar delays, 1.7-1.8 seconds (mean 2.0-2.3). We performed a two-way ANOVA controlling for participant, using transformed delay as dependent variable, and memory task and Twitch activity as independent variables. The ANOVA found a significant main effect of Twitch activity ($F(6, 5378)=25.8, p <.001$). Post-hoc Tukey comparisons found no significant differences in delay between any Twitch activities and the slide-to-unlock control (all $p >.55$). All activities, including the slide-to-unlock control, were significantly slower than the previous task being another back task (all $p <.001$).

We compared accuracy across conditions using a logistic regression, again controlling for participant. Memory task was significant ($\beta = -0.45, z = -4.3, p <.001$), indicating that 3-back tasks produced less accurate performance than 2-back tasks. However, none of the Twitch activities or the control slide-to-unlock activity had a significant impact on accuracy.

Thus, contrary to our hypothesis, Twitch activities had no measurable impact on accuracy or delay for the working memory task compared to a standard slide-to-unlock gesture. While unexpected, this is good news for Twitch: Twitch activities may not meaningfully distract the user from their daily activities while using their phone.

2.6 Discussion

Twitch crowdsourcing can extend beyond the three applications outlined in this paper: opportunities include citizen science, accessibility, education, and scientific discovery. Users might participate in micro-tutoring sessions by marking an English-language sentence as correct or incorrect to help language learners from the developing world and K-12 education. They might answer short survey questions (e.g., txtEagle and Jana [65]), help power accessible interfaces for other users, collate local news, or direct lost people in their neighborhood. Of course, while we have focused on volunteer crowdsourcing, it may be possible to add a microtask marketplace to Twitch so that users can make money as they go about their day.

Longevity and retention are challenges for every crowdsourcing campaign. In our deployment, the median Twitch user was active for one month. (One month out of a forty-day field deployment period is longer than we expected for users of research software.) Moving forward, we believe that the key to long-term usage is finding design opportunities that match users intrinsic motivation. Similar to gamification, twitch crowdsourcing cannot make an uninteresting application more attractive, but it can empower an already viable design. Our results indicated that users would be more likely to continue using the application if they see their progress and value toward the larger goal. In response, we have been developing far richer feedback for the Structuring the Web application (Figure 2.5).

In our vision of twitch crowdsourcing, tasks are fast and impose little cognitive

load. However, many worthwhile crowdsourcing goals may not easily fit this definition. Structuring the Web is nearing this limit, taking roughly three seconds per unlock. We believe that it may be possible to break longer task into smaller ones that can be completed serially across multiple screens. This might enable a user to execute a long task through multiple unlocks, with each fragment taking less than 2 seconds.

Reciprocally, some users have slightly longer stretches of time to contribute, for example while waiting for a meeting to start. Twitch crowdsourcing could apply in these situations as well, for example by allowing users to launch a standalone application or ask the unlock screen to keep cycling through new tasks without disappearing. To maintain users desired level of security, Twitch allows users to add its unlock screen after an Android native unlock. The effect is extremely similar to the passcode unlock on the Apple iPhone: users must both swipe to unlock and enter a passcode. Moving forward, we believe that it would be possible to create Twitch tasks that could function as security locks as well. One possible approach would be to integrate a user-determined security swipe pattern into the motion needed to complete a Twitch task. For example, users could tap to indicate their answer, and then swipe a gesture password from that starting position to verify their identity.

Our deployment underscored that a stream of fresh tasks is required to keep Twitch interesting. This is possible with applications like Structuring the Web, but Census suffered from boredom in repeat usage. The system could be extended to assign microtasks more intelligently. For example, location-based tasks should be distributed among nearby users and topic-based text extractions and other tasks should be dis-

tributed among experts and enthusiasts in each field. Private groups could split up and share their own tasks among friends and coworkers. Likewise, the system could learn to regularly swap stale tasks for fresh ones.

2.7 Conclusion

This chapter presents twitch crowdsourcing, an approach to crowdsourcing based on quick microtasks that encourage short bursts of contribution. Our Twitch application augments the lock screen on mobile phones to harness users' spare moments for short tasks. Unlocking a mobile device via a Twitch microtask takes tenths of a second longer than a standard unlock so little that the two were not statistically distinguishable in our field experiment. Our study, designed to determine the mental workload imposed by Twitch, revealed that our twitch crowdsourcing tasks likewise do not impose a more significant cognitive load than a simple swipe-to-unlock motion. In a public deployment, Twitch collected over eleven thousand data points to help build a census of local activity, rank stock photos, and structure the web.

Reducing concerns about effort, time, and motivation are key steps to greater volunteerism in crowdsourcing. We suggest that small contributions made in short bursts during spare seconds can be aggregated to accomplish significant, customized tasks without placing an undue burden on volunteers. We envision that this approach could bring experts into the crowdsourcing fold, overcoming their historical tendency to stay out because of large time commitments. If we succeed in involving a broader set of

participants and topic experts, they could unlock many new opportunities for research and practice in crowdsourcing.

Chapter 3

The Whodunit Challenge: Mobilizing the Crowd in India

3.1 Introduction

Recent years have witnessed the power of crowdsourcing as a tool for solving important societal challenges [130, 74, 51, 55]. Of particular note are instances of crowd mobilization, where large groups of people work together in service of a common goal. A landmark demonstration of crowd mobilization is the DARPA Network Challenge, where teams competed to find 10 red balloons that were hidden across the United States [7]. The winning team found all the balloons in less than nine hours, utilizing a recursive incentive structure that rewarded participants both for joining the search as well as for growing the team [103]. Since then, mobilization exercises such as the Tag Challenge have shown that teams can locate people of interest across North

America and Europe [108]. The MyHeartMap Challenge mapped over 1,500 defibrillators in Philadelphia County [26]. Authorities have also turned to crowd mobilization for help gathering intelligence surrounding the London riots [29] and the Boston Marathon bombings [6], though the results have not been without pitfalls [3] and controversy [42].

One limitation of prior crowd mobilization studies is that they have focused exclusively on North America and Europe, where Internet penetration is so high that most teams pursue purely online strategies. However, in other areas of the world, the Internet remains only one of several complementary channels for effective mobilization of crowd. For example, in India, 1.2% of households have broadband Internet access [35], but there are 929 million mobile subscribers, over 550 million viewers of television, and over 160 million listeners to radio [35, 4]. An SMS-based social network called SMS GupShup has 66 million subscribers in India [14]. Moreover, there is a rich oral tradition of conveying stories and information face-to-face. Environments such as the Indian railways serving 175 million passengers every week [16] provide fertile grounds for mobilizing crowds. India also has a unique social milieu, with its own social hierarchies, attitudes towards privacy [84], and trust in / responsiveness to various incentive schemes. In light of all these characteristics, it stands to reason that effective crowd mobilization in India would require broader and more inclusive techniques than in Western contexts.

To further explore the landscape of crowd mobilization in India, this paper reports on a new mobilization contest that was designed specifically for the Indian context. Dubbed the Whodunit Challenge, the contest enabled participation through mobile phones instead of via the Internet. The contest offered a Rs. 100,000 (USD

1,667) prize for solving a fictional mystery case, in which teams were asked to gather five pieces of information: Who, What, Where, When, and Why. To participate, an individual had to send a missed call [103] to the contest phone number, which returned via SMS one of five phrases, each providing one of the pieces of information. Because some phrases were returned with low probability, and only one phrase was sent to each phone number irrespective of the number of missed calls received, participants needed to form teams of several hundred people in order to have a chance of winning. In this paper, we use an exchange rate of 1 USD = Rs. 60.

The Whodunit Challenge attracted over 7,700 participants within the first day, and was won by a university team in just over five hours. To understand teams experiences and strategies, we conducted 84 phone interviews, covering most individuals who submitted 3 or more phrases or who received phrases sent with low probability. While many of the winning teams did utilize the Internet to mobilize the crowd for finding phrases, we also uncovered interesting cases that relied mainly on face-to-face or mobile communication. Unlike previous crowd mobilization challenges, many successful teams relied only on personal networks, rather than trying to incentivize strangers to help them search for phrases. Members of these teams were usually unaware of (or unmotivated by) the cash award.

In the remainder of this chapter, we describe the design rationale, execution strategy, and detailed evaluation of the Whodunit Challenge. To the best of our knowledge, this is the first paper to describe a large-scale crowd mobilization contest in a developing country context, exploring the portfolio of online and offline communication

strategies that teams employed. We also offer recommendations to inform the design of future crowd mobilization challenges targeting low-income environments in developing countries.

3.2 Related Work

There is a vibrant conversation in the research community surrounding the future of crowd work [82]. Research that is most closely related to our work falls in two areas: crowd mobilization challenges and crowdsourcing in developing regions.

One of the most high-profile experiments in crowd mobilization was DARPA’s Network Challenge, launched in 2009. By asking teams to find ten red balloons that were hidden across the United States, the challenge aimed to explore the power of the Internet and social networks in mobilizing large groups to solve difficult, time-critical problems [7]. The winning team, from MIT, located all of the balloons within nine hours [122] using a recursive incentive mechanism that rewarded people for reporting balloons and for recruiting others to look for balloons [103]. This approach was inspired by the work of Dodds et al. [63], which emphasizes the importance of individual financial incentives [92]. Cebrian and colleagues proved that MIT’s incentive scheme is optimal in terms of minimizing the investment to recover information [54], and that it is robust to misinformation [95].

The DARPA Network Challenge seeded broad interest in the role of social net-works in homeland security [68]. This led to a follow-up contest called the Tag

Challenge from the U.S. Department of State [108], in which the task was to find five people across five cities and two continents within twelve hours [114]. The winning team found three of the five people and used an incentive scheme similar to the one that won the Network Challenge. Private firms and universities have also explored the potential of crowd mobilization. In 2009, Wired Magazine launched the Vanish Challenge [41] and in 2012, the University of Pennsylvania launched the MyHeartMap Challenge. The latter challenge saw over 300 participants who found and catalogued over 1,500 defibrillators in Philadelphia County [26]. However, to the best of our knowledge, there has not yet been any social mobilization contest with a focus on a developing country. There is a need to explore the landscape of crowd mobilization in developing countries and to identify the differences from crowd mobilization strategies observed in the developed world.

Researchers have also studied the potential and limitations of crowdsourcing in developing regions. Platforms such as txtEagle [65] and mClerk [72] aim to enable workers to earn supplemental income on low-end mobile phones. Others have examined the usage [102, 111] and non-usage [80] of Mechanical Turk in India, where approximately one third of Turkers reside. Efforts such as Ushahidi [71] and Mission 4636 in Haiti [76] have leveraged crowd workers to respond to crises in developing countries. Researchers have also explored the role of social networks such as Facebook [133] and SMS GupShup [109] in low-income environments.

3.3 The Whodunit Challenge

The Whodunit Challenge was an India-wide social mobilization contest that awarded 100,000 Rupees (USD 1,667) to the winner. The objective of the challenge was to understand mechanisms, incentives and mediums people in India use to mobilize large groups of people for a time-bounded task.

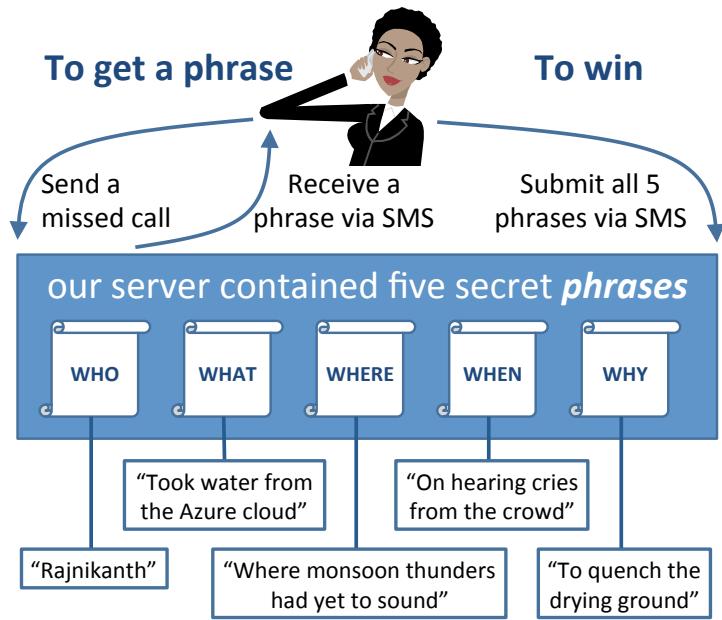


Figure 3.1: Graphical illustration of the Whodunit Challenge.

3.3.1 Design Principles

The Whodunit Challenge embodied three design principles to make it broadly accessible throughout India. In India, 72% of the adult population is illiterate in En-

glish [62]. Thus, we localized the SMS messages by translating them into ten regional languages of India, making them more accessible than contests based on English alone. To ensure that the messages were not distorted in the translation, the translations were done by native speakers of local languages who were highly skilled in English. A majority of the Indian population has constrained access to modern devices and net-works: the smartphone penetration is only 10% [93] and Internet penetration is 20% [17]. Thus, we aimed to enable participation by owners of basic mobile phones, thereby ruling out any dependence on computers, smart phones, or Internet connections (broadband or mobile). While Internet access could still offer advantages to participants, it was not strictly necessary to compete and win. Around 60% of the Indian population earns less than US\$2 per day [30]. Thus, we aimed to minimize the costs of participation. To participate in the contest, users needed to send a missed call from a mobile phone (which incurs no cost to them). To submit a phrase, they needed to send an SMS; this costs at most US\$0.015, though is free under many mobile subscription plans. Our design did not require users to initiate any voice calls, as this expense could have thwarted participation from cost-sensitive groups.

3.3.2 Contest Mechanics

The challenge required participants to reconstruct a secret sentence consisting of five pieces of information Who, What, Where, When and Why (see Figure 3.1). Each piece of information was referred to as a phrase and represented a part of the secret sentence.

To receive a phrase, participants simply sent a missed call to the contest phone number. On receiving the call, our server responded with an SMS containing one of the five phrases. Each phrase was sent in two languages: English and the predominant local language in the telecom circle from which the call was made. The first person to forward all five phrases (i.e., the secret sentence) to our server via SMS was declared the winner. User responses were passed through a transliteration API, providing robustness to any minor typos incurred in re-typing phrases.

What made the challenge difficult is that some phrases were very rare, thereby requiring participants to form large teams to gather all the phrases. Also, we made it difficult for any one person to receive many phrases by sending only a single phrase to each phone number even if we received multiple missed calls from the same number. Regulations in India make it difficult for a person to obtain many phone numbers; for example, VoIP DID numbers are not available for sale (and our server ignored VoIP calls anyway). Also, telecom operators offer a limited number of SIMs per customer, and each requires several pages of paperwork and supporting documents (personal identification, proof of address, etc.). While we advised participants that a very large team would be necessary to win, the award itself was made to an individual. Thus, any sharing of the award within a team would need to be managed by a team leader.

While the Whodunit Challenge was framed in lighthearted terms, we intended for the search for phrases to closely mirror the search for serious time-sensitive information, such as missing persons, suspicious containers, counterfeit currencies, etc. By using electronic phrases instead of physical artifacts, we were able to monitor and control each

step of the contest.

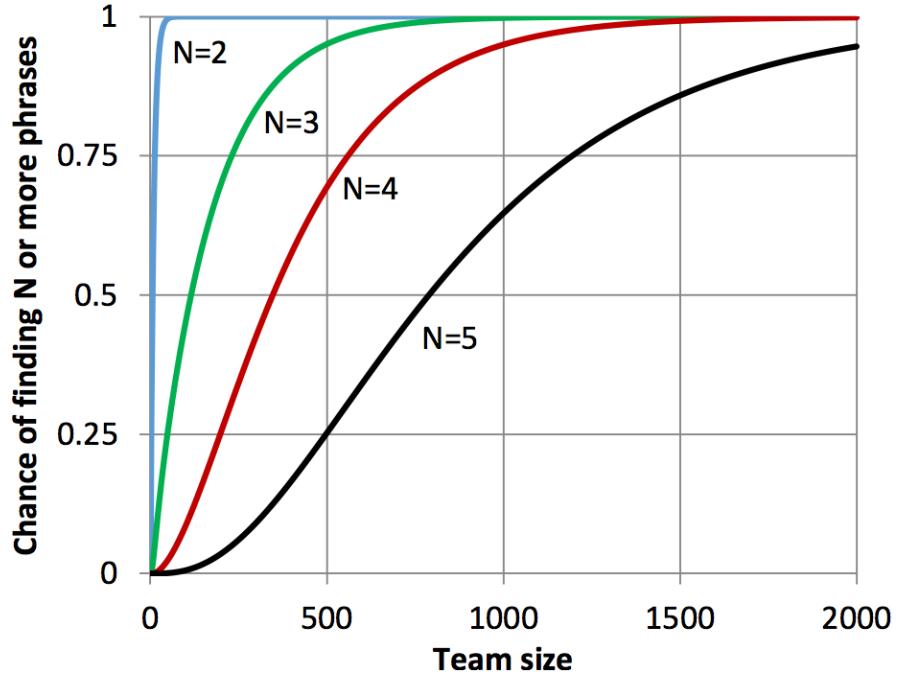


Figure 3.2: Chance of finding N or more phrases as a function of team size.

3.3.3 Chance of Winning

How large of a team was needed in order to win the challenge? We did not publicize this information broadly, though during one Q&A session, we indicated that competitive teams would contain several hundred members. In response to each missed call, the server responded according to a weighted random function, returning *Who*, *What*, *Where*, *When* and *Why* with probability 89.4%, 10%, 0.2%, 0.2%, and 0.2%, respectively. Given these probabilities, the chance of winning as a function of team size is illustrated in Figure 3.2. To have a 50% chance of winning, a team needed 789

people. However, depending on their luck, smaller or larger teams could also win. To have a 5% chance of winning, a team needed about 230 people; for a 95% chance of winning, a team needed about 2040 people. The probability of winning did not depend on participants location, time of sending a missed call, or other factors, as each phrase was returned independently at random.

3.3.4 Publicity and Outreach

We publicized the challenge widely in order to seed participation. A distinguished speaker announced the challenge to a live audience of 2,500 undergraduate engineering students about one week prior to the contest launch [10]. We conducted a large email and social media campaign targeting engineering colleges, MBA colleges, and student volunteers connected with Microsoft Research India. We also presented posters at two academic conferences in the month preceding the contest to create awareness among computer scientists. While the audiences for these activities were primarily composed of Internet users, we advised team leaders that outreach to non-Internet users would be highly advantageous for growing a large team and winning the challenge. Also, to seed visibility among non-Internet users, we met with a group of cab drivers and called ten group owners on SMS GupShup. Our outreach activities led to media coverage by both domestic and international outlets [23, 33]. The basic rules for the contest were explained in the digital promotional material and personal conversations. Internet users could also visit the contest website [37] for more detailed examples.

3.4 Analysis Methodology

To understand the results of the challenge, we employed a mix of quantitative and qualitative methods. We kept electronic logs of all calls and SMSs submitted to our server, and analyzed the approximate geographic origin of calls using the prefix of the telephone number [93]. On the qualitative side, we conducted structured phone inter-views with 84 participants, probing themes such as how they came to learn about the challenge, who they told and how they communicated about it, and what was their strategy (if any) to win. The interviews were conducted in English and Hindi by the first author (male, age 28). Each phone interview lasted around 15 minutes. We took detailed notes during the interview and used open coding to analyze the data. Of the 84 people we interviewed, 65 were students, 17 were employed in a private job, and 2 were homemakers. The specific participants interviewed were 31 people (of 32 participants) who submitted all five phrases; 1 person (out of 2) who submitted 4 phrases; 6 people (out of 6) who submitted 3 phrases; 38 people (out of 53) who received one of the rare phrases (where, when, or why); and 8 other participants.

At the end of the challenge, we also invited participants to complete a brief online survey. We publicized the survey via SMS and also on the contest website, and received about 300 responses in one day. Many questions in the survey were optional and thus, different questions were answered by a different number of users. There were 167 male and 46 female respondents. The average age of the respondents was 21.4 years (s.d.= 6.28). The respondents were from 42 universities and 5 organizations. Respon-

dents included 174 students, 14 salaried employees, 2 professors, and 1 homemaker. The majority of the users had a feature phone or basic phone. Fifty-nine respondents heard about the challenge through an email sent by a friend, college authorities or professors, 58 heard through offline conversations with friends, relatives, professors and colleagues, 47 got the information through Facebook and websites, and the remainder heard about the challenge through text messages, offline promotional events, advertisements, and tasks on Amazon Mechanical Turk. Most respondents, 192, received *Who*, 27 received *What*, 4 received *Where*, 2 received *Why* and none received *When*. Sixty-one respondents reported discovering one phrase while 65, 24, 11 and 36 participants reported discovering two, three, four and five phrases respectively. Eleven respondents could not even begin their campaign as the challenge finished much earlier than they expected. On an average, each person reported sharing their phrase with 33 people (s.d.=120) and receiving a phrase from 30 people (s.d.=93).

3.5 Results

The Whodunit Challenge was launched on February 1, 2013 at 9:00 AM local time. The challenge drew 7,739 participants in less than 15 hours (see Figure 3.3). The first winning submission was made in just over 5 hours. However, we delayed announcing that the contest was over until the evening, as we also wanted to rank and recognize the runner-up teams.

Participants sent a total of 10,577 missed calls to the system. Of the unique

callers, 6,980 received the phrase for “Who”; 740 received “What”; 18 received “Where”, 17 received “When” and 17 received “Why”.

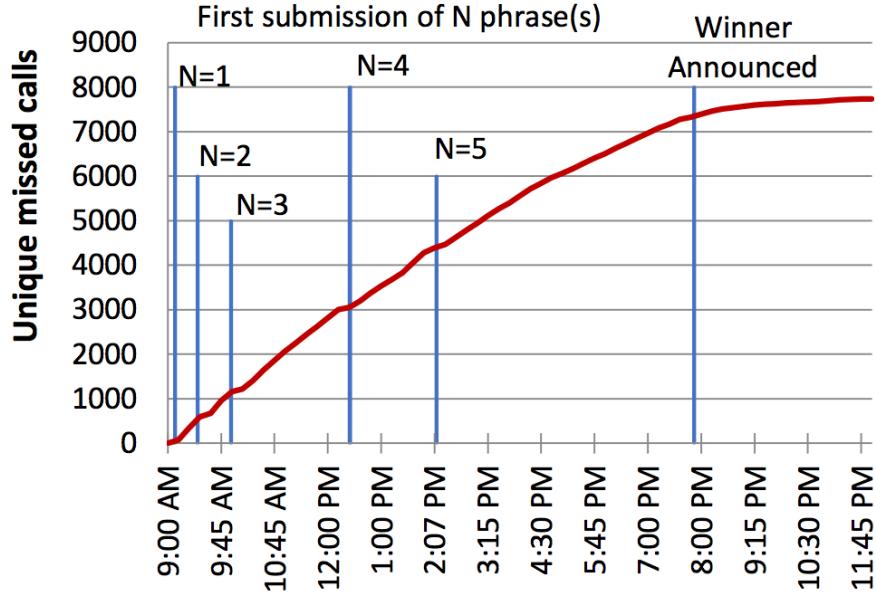


Figure 3.3: Number of unique missed calls vs. time.

There were 185 people who submitted at least one phrase. The first person to submit two phrases did so within 26 minutes; 3 phrases, within 57 minutes; 4 phrases, within 3 hours and 19 minutes; and five phrases (winning the contest) after 5 hours and 7 minutes. Geographically, participation spanned across all of India, as illustrated in Figure 3.4.

3.5.1 Winning Strategies

The winning teams are listed in Figure 3.5. The figure lists all 20 teams who submitted 3 or more phrases and, to the best of our knowledge, discovered these phrases

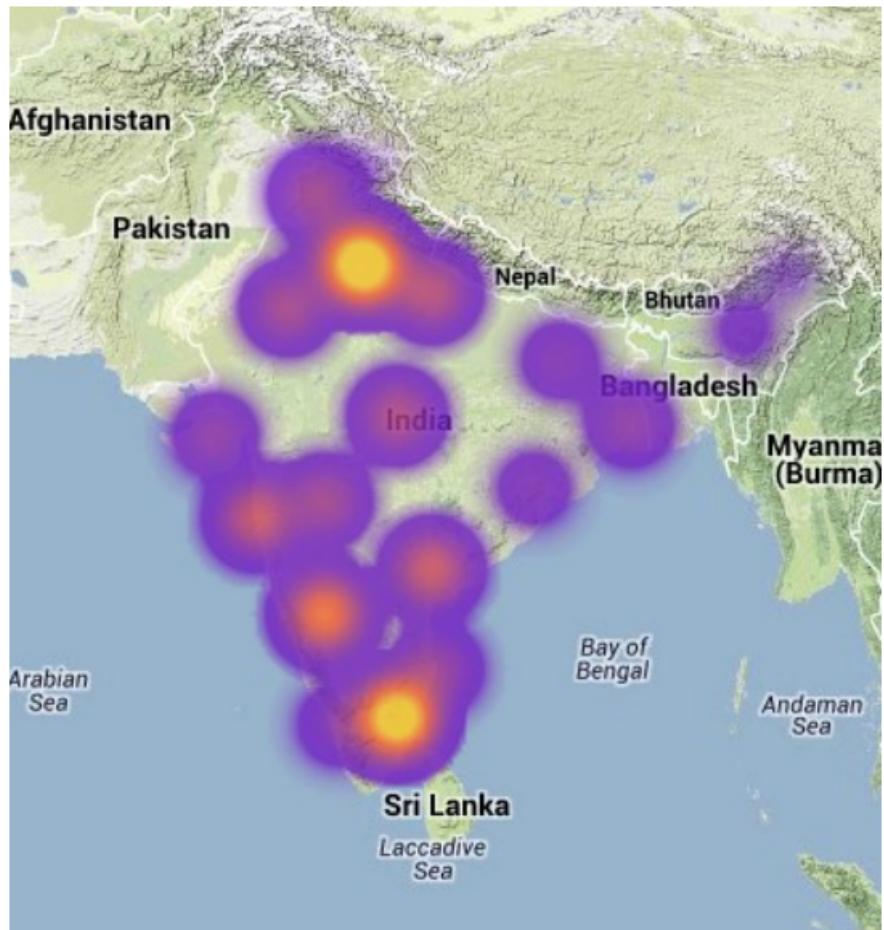


Figure 3.4: Heat map of received missed calls.

without help from other teams. While we are certain about the rank ordering of the first two teams, there is a complication in ranking the remaining teams: the winning team posted all of the phrases on the Facebook page of Whodunit Challenge at 4:30pm. Thus, we rank teams by two criteria: first, by the number of phrases they submitted in advance of 4:30pm, and second, by the total number of phrases they submitted and claimed (during our interview) to have found independently. While 13 teams claimed

to have found all the phrases on their own, only 2 teams found all phrases in advance of the leak.

Team Number	Affiliation	Phrases submitted by 4:30pm	Total Phrases Submitted *	Time submitted last phrase	Face-to-face	Voice Call	SMS	WhatsApp	Email	Social media	Benefactors of Prize	Notes
1	IIIT Delhi (1)	5	5	2:07 PM	✓	✓	✓			✓	published incentive scheme	used SMS server; Facebook group of 474
2	IIT Delhi (1)	5	5	2:14 PM	✓	✓				✓	team leaders only	mostly used voice, SMS to reach to friends & family
3	IIT Delhi (2)	4	5	5:00 PM						✓	published incentive scheme (see text)	website with 200 registrations; FB event with 392 replies
4	Jansons Inst. of Tech.	4	5	7:00 PM	✓	✓				✓	shared with team (details unclear)	50% reached via SMS, voice; 50% via FB
5	Paavai Eng. College	4	4	3:30 PM	✓	✓	✓			✓	team leaders only	2 leaders managed 7 sub-teams of 15-20 each
6	IIIT Delhi (2)	3	5	7:05 PM	✓						team leaders only	leaders focused on different geographies
7	IIIT Delhi (3)	3	3	2:10 PM	✓	✓	✓	✓	✓	✓	\$180-\$270 for reporting new phrase	one-person team; calls & WhatsApp worked best
8	IIM Indore	3	3	2:24 PM	✓	✓	✓				given to leaders, who distribute to sub-teams	focused on calls & SMS

9	Delhi Univ.	3	3	3:18 PM		✓	✓			✓	team leaders and out-of-state champions	focused on calls, as many do not read SMS
10	VIT Chennai	2	5	7:07 PM			✓				mostly leaders; small share (TBD) with team	used SMS exclusively
11	UPEI	2	3	5:11 PM	✓		✓	✓			team leader only	one-person team
12	LBS Institute	2	3	5:43 PM	✓						donate to college	team leaders were classmates
13	MIT Manipal	0	5	4:59 PM				✓			team leaders only	relatives in hometown spread info to many
14	Chandigarh	0	5	5:45 PM	✓	✓	✓				team leaders only	mother/daughter team; reached to friends/fam
15	IIM Ahmedabad	0	5	6:01 PM	✓				✓		team leaders only	had classmates make two calls: local/home SIM
16	Class 11 students	0	5	6:54 PM			✓				team leaders only	main team leader is junior in high school
17	Amrita School of Engineering	0	5	7:00 PM				✓		✓	sponsor industrial visit for college	leaders asked friends to contact friends at home
18	VIT Chennai (2)	0	5	7:48 PM		✓	✓			✓	promised party for team	made voice calls to explain contest purpose
19	VIT Chennai (3)	0	5	7:54 PM		✓	✓			✓	promised party for team	70% reached via FB; 30% via calls and SMS
20	Unknown	0	4	5:32 PM	‡	‡	‡	‡	‡	‡		‡

Figure 3.5: Top 20 teams in the Whodunit Challenge. (*) We asked teams to report the total number of phrases that they submitted without help from. (‡) Data not available.

The winning team was based at the Indraprastha Institute of Information

Technology Delhi (IIIT Delhi), led by 2 Ph.D. students and 6 undergraduates. In advance of the contest launch, this team set up a website and a Facebook group that attracted 474 members. The website publicized the following financial incentives. If the team won, they would award Rs. 10,000 (USD 167) to anyone who sent them a new phrase; Rs. 2,000 (USD 33) to anyone who directly referred someone who sent a new phrase, and a mobile top-up worth Rs. 50 (USD 0.83) to the first 200 people who sent any phrase. They set up an SMS server to which people could forward phrases. They recruited team members using a variety of methods, spanning phone calls, SMS, WhatsApp and social media platforms.

The second-place team was based at the Indian Institute of Technology Delhi (IIT Delhi), led by eight second year Computer Science undergraduates. This team finished just 7 minutes behind the leader. Yet they used a very different strategy: they set up a small call center, relying mostly on direct calls and SMS to reach out to family and friends who live in smaller towns and villages across the country. In turn, they asked these contacts to gather team members from the local community. One team member also set up a Facebook group and utilized Facebook group chat.

The third-place team was also based at IIT Delhi, led by six undergraduate students. This team found 4 phrases in advance of 4:30pm, and claims to have found the fifth phrase (working independently) by 5:00pm. Unlike other teams, this team relied solely on social media and email to recruit members. They invited over 4,000 people to a Facebook event, of whom 329 replied with Going and 63 replied with May-be. The group page was linked to another website where team members could register and

receive a unique ID, which could be used to refer others to the team. Participation by those referred led to modest payments to the referrer (Rs. 100, or USD 1.67, for 20 referrals).

The fourth-place team, based at the Jansons Institute of Technology in Coimbatore, was led by a single undergraduate student. She estimated that she reached out to 250-300 people, half via SMS and voice calls, and half via Facebook. She submitted the fourth phrase at 3:30pm and the fifth at 7:00pm. While she expressed interest in sharing the prize money with team members, she did not have any incentive structure in place and the terms were not discussed with the team members; her team members helped her as a personal favor rather than for a monetary incentive.

The fifth-place team was based at Paavai Engineering College in Tamil Nadu, led by two cousins. They managed seven sub-teams with 15-20 people per team and used face-to-face interactions, phone calls, SMS, and social networks to coordinate. Interestingly, they also contacted a relative who worked at a mobile shop; the shop asked customers to give a missed call on the contest number and forward phrases to him, which he then shared with the team leaders. They did not have a formal incentive strategy, though as they got closer to winning, they offered to share a prize with those who helped them.

3.5.2 Emergent Themes

Rather than describe additional teams in detail, we present three high-level themes that emerged across the remainder of our analysis. This draws from our in-

terviews with teams, our interviews with recipients of rare phrases, and the web-based follow-up survey.

3.5.2.1 Internet but also SMS, Voice, Face-to-Face

All of the top five teams (and 14 of the top 19) utilized the Internet to their advantage. The most common uses were to establish a website (either independently or as a Facebook page) and to reach out to friends and contacts via Facebook (10 teams), WhatsApp (6 teams) and email (2 teams).

At the same time, teams also demonstrated a heavy reliance on non-Internet technologies: thirteen teams utilized SMS and eleven utilized voice calls to mobilize people. There were nine teams that utilized all three technologies: SMS, voice calls, and the Internet. Only three teams relied on Internet technologies alone.

The prevalence of communications outside the Internet is confirmed by our interviews with those who received rare phrases. Most often, they heard about the contest via an SMS ($n=9$) or face-to-face interaction ($n=8$) with a friend. Learning about the contest from Facebook was less common ($n=4$). Other ways of learning about the contest included email, phone calls, WhatsApp, etc. Our online survey revealed that even among participants that have access to various Internet-based services, 43% heard about the contest through offline personal conversations with friends and colleagues.

An example of effective use of non-Internet technologies is the runner-up team (IIT Delhi), who relied mainly on a call center approach to reach family members in rural India. As a team leader explained to us, “My mom doesn't use Internet”, and

neither does the majority of Indias rural population, which constitutes 72% of the overall population. As another example, a team of office drivers used only voice calls to manage a team and found two phrases in less than two hours.

One enterprising undergraduate (from the Amrita School of Engineering, Coimbatore) claims to have built a team of 200 peers using face-to-face contact alone. He estimates that with the help of his team's combined efforts, he reached out to at least 1,000 people. While he reports finding three phrases, he did not submit them via SMS because he thought more phrases would be released later. (For this reason, his team does not appear in Figure 3.5.).

These teams success illustrates that it is also possible to mobilize a sizable crowd without broadcast technologies such as social media or Internet websites. This finding has implications for social mobilization in areas lacking Internet connectivity, or for well-connected areas that experience Internet outages during crises.

3.5.2.2 Reliance on Personal Networks

In previous social mobilization contests, many teams incentivized strangers to join them. However, in the Whodunit challenge, a common thread amongst many teams strategies was a reliance on personal networks: team leaders reached out to their friends and family, as opposed to incentivizing lesser-known acquaintances or strangers to join their team. This is already evident in the strategies for teams 2, 4, and 5, described above, as well as for many other teams. This trend is also corroborated by the online survey where 63 respondents reported relying on friends and colleagues for

discovering phrases rather than 16 respondents who incentivized strangers. We also collected anecdotes where participants simply borrowed their friends phones and gave the missed call on their behalf, without even explaining that there was a contest. If a new phrase was received on the friends phone, it would be forwarded to the participants phone. Three recipients of rare phrases reported that their phone was borrowed and used in this way. One recipient of a rare phrase was a vegetable seller who had absolutely no knowledge about the contest; we hypothesize that his phone was borrowed without offering him any explanation.

3.5.2.3 Most Participants not Driven by Cash Rewards

Building on the prior theme, the primary motivation for most participants was a desire to help a friend or family member, rather than any desire for (or even knowledge about) a cash award. Of the top 19 teams, less than half had any plans to distribute the cash prize beyond the inner circle of team leaders; even the runner-up team did not offer any financial incentive to its members. In teams that did plan to distribute the prize, the majority were very vague about how they might reward their full team. In contrast to the challenges conducted in developed countries, team members were motivated by non-financial factors, and any reward offered to them would be perceived more as a courtesy than as an owed compensation for their services.

We can quantify this tendency based on our interviews with those who received a rare phrase. Of the 35 respondents, only about one quarter (9) said that they were told about any financial incentive in relation to the contest. The majority (18) were

not told about incentives, while the remainder (8) were team leaders or individuals who were working alone. Of the people who were not told about any incentive scheme, the majority (12/18) nonetheless shared their phrase with others. This fraction is not significantly different from those who shared their phrase with knowledge of an incentive scheme (7/9). Some team leaders offered non-monetary incentives for their members. The leader of a team from Amrita School of Engineering, Coimbatore (#17 in Figure 3.5) promised to sponsor a forthcoming industrial visit for his class if they won the challenge. We talked to four team leaders who proposed to throw a party for their friends if they were the winner.

Participants sometimes had intrinsic motivation to participate. For example, one student who received a rare phrase and forwarded it to benefit the winning team (IIIT Delhi) remarked, I knew about the incentive model. Money was not important. I wanted my institute to win. An Infosys employee chose to participate and forward a rare phrase to a friends team because he thought the contest itself was creative and worthy of participation. He thought that the purpose of the contest was to understand the role of technology in solving crime.

Some teams also experimented with other incentives. One team (not shown in Figure 3.5) approached a charitable foundation, asking them to help publicize their team in exchange for 75% of the prize money. While the foundation was receptive, it did not promptly post about the challenge on its Facebook page (which has over 20,000 likes) and thus offered little of the anticipated help within the timespan of the contest.

3.6 Discussion and Recommendations

While the Whodunit Challenge was quite successful in attracting enthusiastic participants from across India, the lessons learned can also serve as design recommendations to help future crowd mobilization challenges to reach out to a larger number of people, especially in low-income or offline environments.

One of the shortcomings of the Whodunit Challenge was the low level of engagement by low-income low-literate populations, primarily because we did not promote the contest widely in offline environments. The contest and the prize money appeared to be too good to be true for many low-income people that we interacted with. Many of them were uncertain about the reasons for awarding a high monetary prize just for sending missed calls. Despite our explanations, they had reservations about whether they would be charged for sending a call to our system. They were also concerned whether we would misuse their number, e.g., by sending them pesky voice calls or text messages.

To encourage more participation by non-Internet users, one approach would be to restrict promotions to offline audiences, limiting the visibility to Internet users. Another approach would be to partner with local organizations that work closely with low-income groups, distribute graphic pamphlets in local languages, and conduct outreach efforts led by people who are from the target community or have similar socioeconomic status. It could also help to make the contest harder, for example, by decreasing the frequency of certain phrases or enforcing geographical diversity of team members (in

India, coarse-grained geographic information can be determined from the caller ID. [24]).

As teams are forced to reach out to broader populations, they may derive greater benefit from reaching out to the masses of rural and lower-connectivity residents. Disseminating phrases in audio format rather than text would also enable inclusion of lesser-educated participants, though the cost of phone calls could be a significant deterrent (either for participants or for the challenge organizers, depending on who pays for the calls.)

One of our interesting findings is that participants were often motivated by non-monetary incentives, including social support for friends and recognition for their institution. Future challenges might employ non-monetary incentives to increase participation, for example, by offering recognition, goods or services that cater to groups (such as a party or travel vacation).

Our usage of mobile phone numbers as a unique personal identifier was largely successful in prompting the formation of large teams. However, it also led to some subtle implications, such as the practice of borrowing others phones to leverage their participation without their full knowledge or consent. While we did not observe any serious abuses of this situation, e.g., by stealing phones or feeding misinformation to potential participants, these possibilities are nonetheless important to consider and guard against in future challenges.

One limitation in the design of the Whodunit Challenge is that it is not possible to know the exact sizes of teams. Addressing this limitation would have required a fundamental change in the contest dynamics, for example, to require each participant to identify themselves with one or more teams. This would likely require an interaction

richer than a missed call, which would have added cost and complexity for participants. Though sending SMS may seem easy, only 185 of 7,739 participants submitted a phrase to our server. Some participants may have been motivated only to share their phrase with their team leader, while other participants may have had limited familiarity with SMS and how to forward them. In any case, finding creative techniques to more accurately track the growth and composition of teams, without adding complexity for participants, could yield large benefits in the analysis of future challenges. One potential approach could be to host a large number of contest phone numbers, each advertised to a small number of people. If two participants place calls on different numbers, it would be unlikely that they are on the same team.

Our final recommendation is to take extra care in designing simple rules and communicating them to participants. Though we distilled the challenge rules to very simple language, including several illustrative examples, many teams misunderstood aspects that prevented them from competing well. We found three teams who thought that some phrases would be released at a later date, preventing them from being aggressive in the initial stages. We talked to five teams who assumed that phrases would be distributed across different geographical regions, causing them to seek out more geographies rather than seeking out more people. We also spoke with five teams who assumed that all phrases needed to be submitted together, preventing them from gaining feedback and recognition for intermediate progress. It is important to anticipate any possible misconceptions and proactively convey the requisite clarifications. Several individuals misunderstood each phrase to be a puzzle instead of a part of the secret

sentence; for example, in response to “*Who: Rajnikanth*”, they would respond with “actor”. While these details are somewhat specific to the Whodunit Challenge, the broader implication is that though it is difficult, it is necessary to design simple rules that are easily understood and easily communicated from one person to another. This is especially important for lesser-educated participants and those who may lack the devices, connectivity or bandwidth to view large explanatory materials (such as websites, promotional videos, etc.). We also recommend setting up more accessible information portals, such as an Interactive Voice Response system, to make the rules more accessible for people with low literacy and limited access to the Internet.

3.7 Conclusion

This chapter presents the first crowd mobilization challenge conducted in India, a developing country context where effective social mobilization is broader and more inclusive than the rich-country settings studied previously. We customized the design of the challenge to incorporate local languages and to enable participation at very low cost by anyone with access to a basic mobile phone. The challenge was successful in attracting broad participation, spanning 7,700 participants from all across India in less than a day. While many participants utilized Internet technologies, we also found interesting usage of SMS, voice, and face-to-face communications that offered benefits in the Indian context. Unlike previous social mobilization contests, participants relied primarily on their personal networks, and often recruited team members with-out offering

any financial incentives. We synthesize our lessons learned as a set of recommendations to help future crowd mobilization challenges extend their reach into low-income, offline environments.

Chapter 4

The Aspiring Researchers Challenge: Crowdsourcing Research at Scale

4.1 Introduction

Scientific research is becoming increasingly collaborative, yet primarily limited to professional researchers in labs and universities. In such a setup, the professional success is subject to highly robust and closed-loop reward structure (eg: publications, grants, tenure) [59]. This traditional approach often limits the range and scale of open ended problems, and access to people interested in solving it. Furthermore, it is constrained by resources like time and funding, therefore restricting high risk and large scale research from happening. Meanwhile, the internet is facilitating large scale creative collaboration - people are getting together to share knowledge (Wikipedia) or write software (Apache, Linux) [132]. This has unleashed a new era for collaboration

and crowdsourcing, where more viewpoints can break traditional assumptions. Building upon these experiences, can we motivate people around the world to come together and do scientific research with experts? Can we provide access and educational value at scale, through massive collaboration and crowdsourcing an open ended problem?

Crowdsourcing an open ended scientific problem is extremely challenging. Besides high coordination and logistical demands, such ventures are limited by the expertise and prior knowledge of the crowd. Therefore, restricting them from making concrete contributions or other pressing commitments can cause them to leave. This inconsistency could make research at scale a risky endeavor, and lower faculty/researcher interest in organizing such initiatives. Furthermore, keeping crowd motivated and giving due credits at scale is unconventional and a non trivial problem. Until now, scientific research has been primarily confined to academia, that follows a consistent structure of project execution and credit distribution. Developing a robust design of the project and planning its execution is extremely important and key to its success. Overall, with highly varying preparation and commitment from crowd, designing the project to coordinate their efforts for high-quality output at scale is extremely hard and challenging.

To address these research questions and challenges, we launched a first of its kind "research-at-scale" initiative - The Aspiring Researchers Challenge (ARC) [38]. As part of the program, we designed a process structure and made an open call to people around the world - ranging from high school students to PhD students and professionals. Over 1,000 participated, from 6 continents; more than 25% of them were women and about 80% had no research experience. The challenge or the program ran three open

ended projects in different areas of computer science - computer vision, human-computer interaction (HCI) and data science. These projects were lead by three professors at Stanford University and University of California, Santa Cruz. Over the span of six months, the work done by crowd produced three work-in-progress papers [70, 117, 128] at top-tier conferences in computer science - ACM UIST and AAAI HCOMP. In the HCI project, a full paper is under review at ACM CHI; while two participants who aspire for graduate school, earned a full time research assistant (RA) position at Stanford University.

In the remainder of this chapter, we describe the coordination and process design, research components, and detailed evaluation of the Aspiring Researchers Challenge (ARC). To the best of our knowledge, this is the first initiative to implement academic style research at scale, targeting an open problem via an open call. This chapter also reflects on the best practices that can inform the design of future initiatives in this space.

4.2 Related Work

As science is becoming more collaborative, there have been attempts at involving people around the world to solve complex scientific problems [31, 79]. Citizen science initiatives are on rise, where projects like GalaxyZoo [107] and NASA Clickworkers [121] involve non-experts to do microtask that lead to novel discoveries. Projects like FoldIt [58] and EteRNA [86] utilize game as a platform and made breakthroughs

in the area of protein folding and RNA sequencing. These approaches utilize focused contributions from the crowd with varying skills to get involved. Yet, there have been few successful examples where people collaborated at scale and made intellectual contributions to do research and produce papers. One notable example is the Polymath project [59], where scientists collaborated online to solve a complex math problem, that does not naturally split up into a vast number of subtasks. This effort brought 39 leading scientists and experts together to build their solutions over each other via a blog. In another endeavor, R Silberzahn et al [118] recruited 29 teams of scientists to analyze same dataset and compare the findings in parallel, leading to a more reliable scientific finding. Though these initiatives embrace research at scale, they depend on collaboration between committed experts in the area, for a very specific research problem, thereby making coordination effort minimal and different from ours.

4.3 The Aspiring Researchers Challenge

A research project goes through several phases and is composed of a few components, that results in a publishable output. Traditionally, a small group of researchers collaborate and work through the brainstorming and paper writing phase. However, given the scale of people involved in the Aspiring Researchers Challenge - it demands a different coordination technique. A technique that can organize people with limited prior knowledge and commitment towards highly intellectual contributions to open ended problems and learning about research topics. This section walks through the

coordination techniques, process and the method.

4.3.1 Coordination Technique

Our project is structured for open ended problems with educational mission, that could be split and involve people irrespective of their prior knowledge or commitment level. To coordinate crowd in a massive open online initiative, we developed a weekly structure that would facilitate collaboration, communication and credit distribution. For a given project, there were primarily four drivers:

1. **Faculty/advisor:** Professors who were leading the project.
2. **Research Assistants (RA):** PhD students who were helping faculty in this endeavor, by managing the process and developing weekly milestones. Each faculty was collaborating with two RAs each.
3. **Crowd/participants:** People who signed up for the program to do research with the faculty.
4. **Directly Responsible Individual (DRI):** Exceptional crowd students who have proven their commitment and willing to volunteer to lead specific goals.

Depending on the phase of research, teams could work in parallel to create divergent ideas in the initial phase [64]. In the later phase, they can form a goal oriented dynamic team to build upon each other's work - fostering collaboration. Our HCI survey ($n=64$) suggested that people enjoyed working in both the phases (median = 4,

in a Likert scale of 5 - where 5 means 'strongly agree'). For efficient functioning, these dynamic teams would be micro managed by DRIs; who would focus to align research goals, self-organize meetings and further coordinate crowd students. This would emulate a multi-tier coordination effort, where crowd organization would be managed by DRIs, DRIs would work with RAs, and RAs work with faculty to help make decisions. In research, it is valuable to have granular level visibility - this approach supports tighter collaboration between all driving forces. Based on a survey in the HCI project, that was filled by 64 people, we asked Likert based questions on a scale of 5. We calculated the median and found that - people liked the DRI model of managing milestones and goals (median=4); people who DRIed also enjoyed managing the micro milestones (median=4); finally, people liked the project coordination approach by the staff (RAs) and professor (median=4).

The goals would be designed by the faculty member and their research assistants, after synthesizing valuable contributions for a given week. These goals would be communicated to crowd via video meetings, followed up by open discussion on a chat platform and forums (as shown in Figure 4.1). The structure is designed to scale, and utilizes peer feedback techniques and badge distribution to converge findings and determine top contributors. To overcome intra-team badging biases, PageRank [101] is applied on the distributed badges. Altogether, these techniques provide educational value to the crowd students, who collaborate to make intellectual contributions and solve open ended problems.

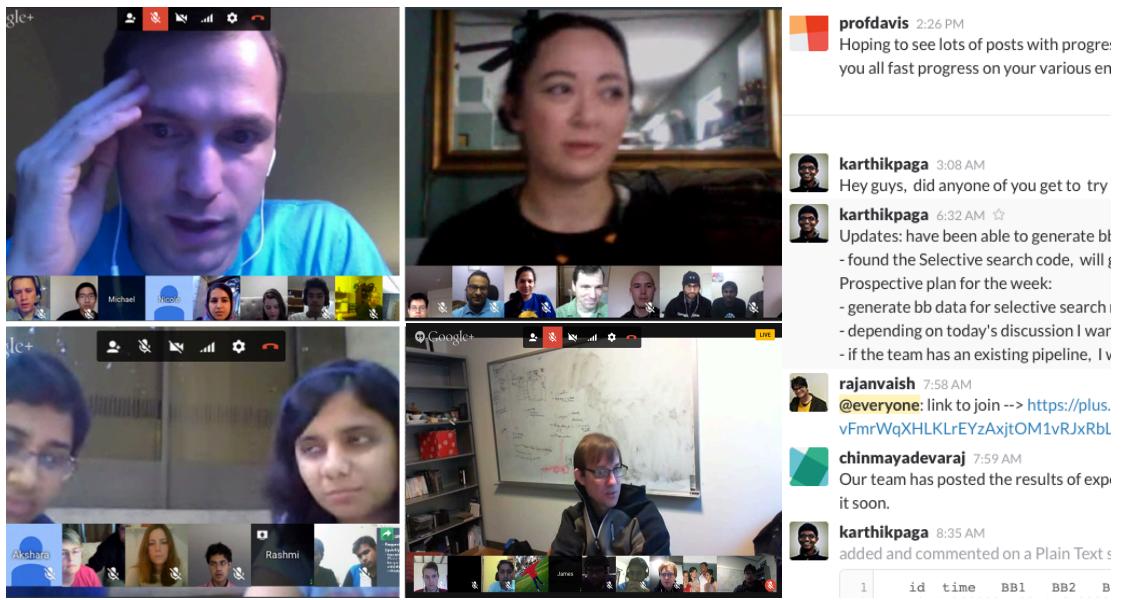


Figure 4.1: Left - Screengrab of various video meeting sessions through Google Hangout on Air; Right - Screengrab of discussions using Slack - a common communication medium.

4.3.2 The Process

On the lines of MOOC, our process began with an open recruitment phase, where we accepted all applicants to maximise the impact. Through our weekly structured approach, we successfully provided scaffolding to train participants with none to minimal research experience. Our structure supported and helped them understand the academic research process as they went through different phases like - brainstorming, prototyping, development, evaluation and paper writing. Though all three projects had variation in their format of execution, they followed a common fundamental approach leading to success.

When divergence is needed to explore multiple solutions and approaches, we

encouraged crowd to work in teams in parallel. For a given goal or topic of concern, crowd participated in discussions on the chat platform (Slack), they self organized meetings, collaboratively wrote ideas and created multiple repositories on Github. This parallelization at scale helped produce the range of ideas and solutions, that is not achievable with a small group of experts. Figure 4.2 shows some of these contributions, that range from storytelling, prototyping, data analysis, brainstorming, voting etc.

To ensure consistency or to make a decision, convergence of ideas and work produced is extremely critical. We encouraged DRIs to lead goal oriented dynamic teams, that were formed to address specific topics. Through the entire week, DRIs would coordinate each effort through focussed micro management and collaborate based on their interest and expertise. These goals were designed by the faculty member and RAs, who would synthesize the most interesting contributions for a given week and discuss next weeks milestones or goals over video chat (Google Hangouts on Air) with the crowd.

To facilitate the convergence process, we encouraged crowd to provide peer feedback on the submissions via upvotes and constructive comments. These comments got reviewed by the faculty and RAs at the end of the week for milestone planning. To infuse consistent motivation and track participation, we integrated credit distribution into the weekly structure. These credits were eventually used to determine author order of the paper written by crowd. The author order further utilized pagerank algorithm to distribute weight based on the contribution, and overcome intra-team badge exchange biases.

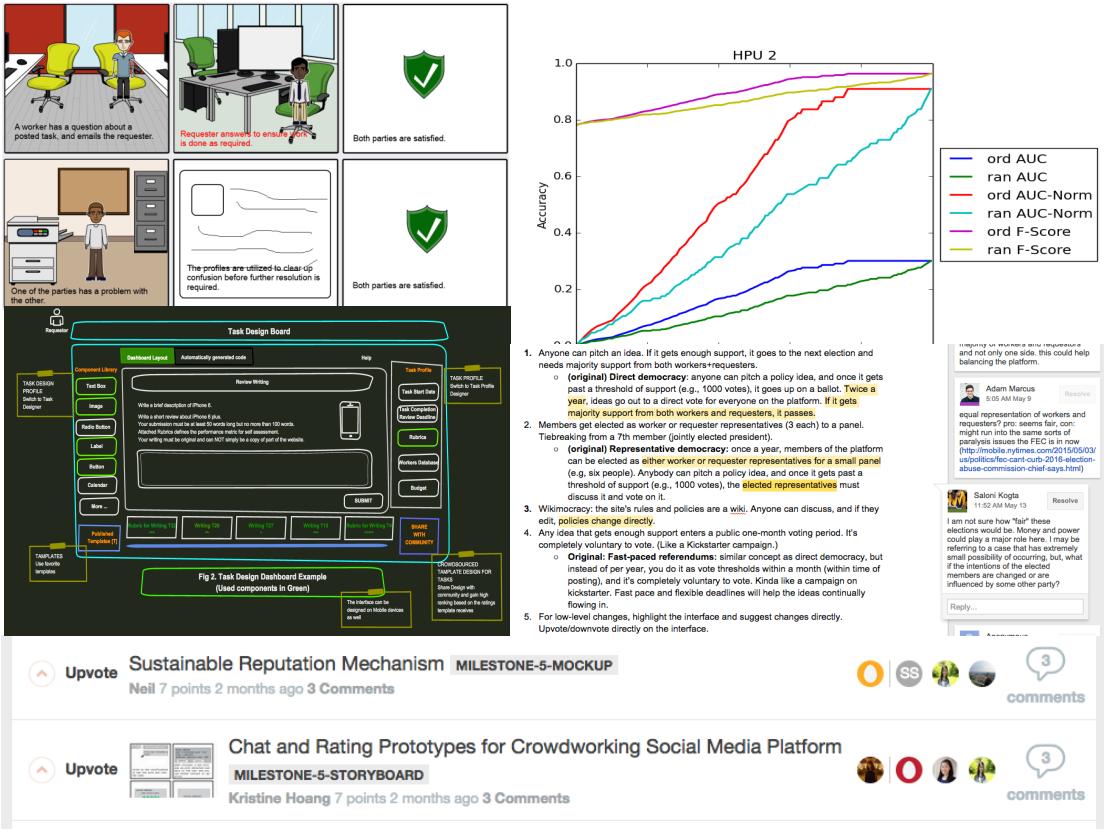


Figure 4.2: Crowd contributed to all the projects in a variety of ways - storytelling, prototyping, data analysis, brainstorming, peer feedback and voting to name a few. Above are a few screengrabs from some of these efforts by crowd.

4.3.3 Projects and Demographics

This weekly structure helped us run three different projects in computer science. The computer vision project was lead by a faculty at UC Santa Cruz, that found strategies for combining computer and human work to reduce cost by as much as 60% [128]. HCI project was lead by a Stanford faculty, that explored and built a self-governed crowdsourcing marketplace (Daemo [34]) - designed to amplify trust in crowd work [70]. Data Science project was lead by another Stanford faculty that explored the

'wisdom of crowd' effect for 1,000 different tasks in 50 subject areas [117]. Figure 4.3

shows the screen grabs of the system built or results found by crowd in these projects.

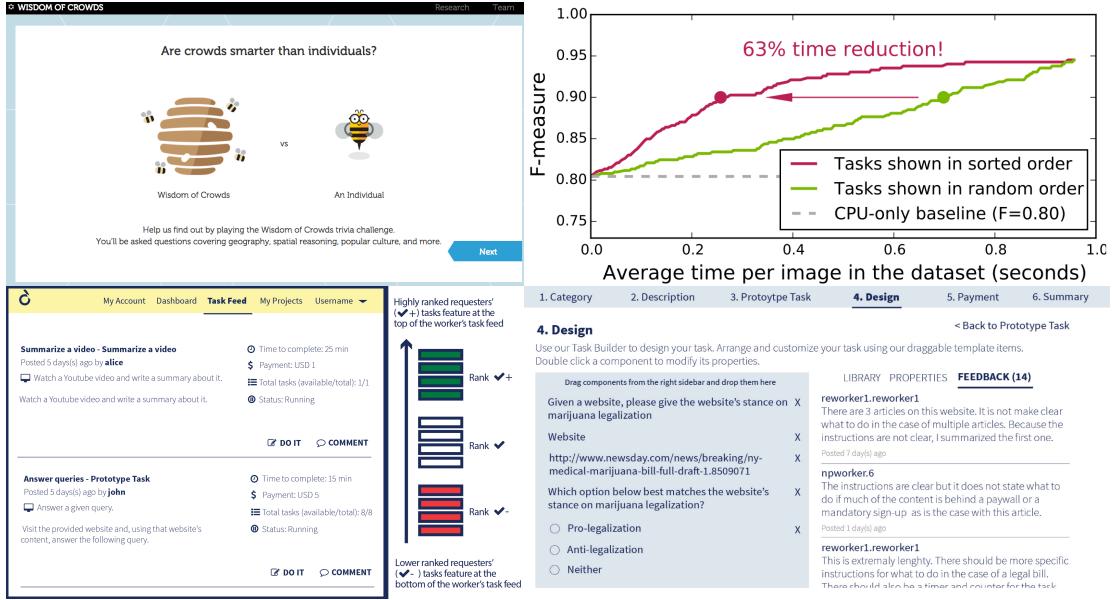


Figure 4.3: Top left - screen grab of wisdomofcrowds.stanford.edu website, that was built as part of the Data Science project. Top right - results from the Computer Vision, showing a cost savings of 63% was achieved in the experiments. Bottom - screen grabs of the crowd platform built by crowd, shows the reputation system and task authoring interface from daemo.stanford.edu.

In collective we recruited 1,097 participants via an open call, representing 6 continents (primarily from India and the United States), where more than 25% of them were women. 259 of them signed up for Computer Vision project, 179 for Data Science, and 659 for HCI. Please note that the HCI project recruited people in two rounds - 286 in first round, 373 people in second round. The median age of our participants was 21 years, who ranged from high school students to professionals in industry. Among our participants, more than 70% were undergraduate students, more than 70% were

pursuing or had a degree in computer science or related areas, and about 80% of the people had never co-authored a research paper before. Projectwise, Computer Vision and Data Science participants were primarily undergraduate students, while HCI project consisted of a diverse set of participants. Our participants were motivated to pursue research with experienced faculty in the area, learn about the process and co-author a paper for their career development.

4.4 Research Components and Findings

Our program was designed to run projects for 12 weeks, imitating a standard internship period or a university quarter. However, it has continued to stretch longer due to crowd interest and participation. Though the Data Science project concluded in 12 weeks, the Computer Vision project continued until 18 weeks and HCI project is still ongoing for more than 30 weeks. During this duration, crowd members have exchanged more than 250,000 messages, watched more than 125,000 minutes of meeting videos, edited the wiki more than 12,000 times and acknowledged about their newly acquired skills to do research. These projects have produced three top tier work-in-progress papers [70, 117, 128] at ACM UIST and AAAI HCOMP, while a full paper on HCI project is under submission at ACM CHI. In the HCI project, the work-in-progress and full paper were primarily written by crowd themselves. The crowd platform built by them successfully completed an assignment from a major enterprise worth \$5,000. Furthermore, two of its top crowd participants who aspire to pursue grad school earned

a full time RA position at Stanford University.

In this first of its kind initiative, there were several factors at play, that lead to these achievements - from team dynamics to communication, from idea evolution to credit distribution. During the course of research, crowd participated and contributed to several of its components - brainstorming, paper writing, author allocation and more. In the following sections, we will talk about many of these factors and their role in the initiative.

4.4.1 Participation and Longevity

A campaign's success highly depends on continued participation by its members. As shown in Figure 4.2, crowd made contributions in several domains. Upon input from 64 people, via an HCI survey - we learned that 81.3% people contributed in brainstorming and discussions, and because each person could have contributed to other aspects of the project - 39.1% did front end coding, while 25% did back end coding. 59.1% helped write the paper, 46.9% helped run studies and evaluation, 51.6% did design and UX, and 17.2% helped with the logistical affairs. This shows that most of the participants made contribution in more than one aspects of the project - making it successful. Meanwhile, it is also valuable to understand their engagement activity, longevity aspects and reasons for dropout. Further in this section we'll reflect on how the participation in our initiative compares with MOOC (massive open online course - another initiative at scale that aims at providing open access to topics of interest to anyone in the world, often taught by university professors online) [25], drop out reasons

and crowd activity on various platforms - Slack [32], Piazza [28], Google Hangout on Air/YouTube [12], Wiki [39] and Github [11].

4.4.1.1 Participation: Communication Activity - Slack

Communication is critical to participation in any voluntary campaign. In our initiative, though Slack was adopted by all the three projects, it was the primary means of communication in the HCI project. Other two primarily utilized Piazza. Over the span of over 30 weeks, the HCI crowd exchanged more than 200,000 messages, primarily via private direct messages (DMs). On average, each week saw 1,620 messages exchanged, and a median of 1,450. In terms of active participants, throughout its duration, the project attracted more than 40 participants every week, averaging to 81.84 with median being 72.

In its first week, the project saw 140 active participants, that was added to 195 during a second round of recruitment in summer (12th week) - see Figure 4.4. The project has consistently seen an active participation by more than 40 participants throughout, with peaks around paper deadlines as can be seen in the Figure 4.4 and Figure 4.5. The crowd also helped write papers, where 61 participants authored UIST work-in-progress paper in the 19th week (68.71% people dropped off, considering 195 participants at peak) and 53 participants helped author CHI full paper in the 30th week (72.82% people dropped off, considering 195 participants at peak).

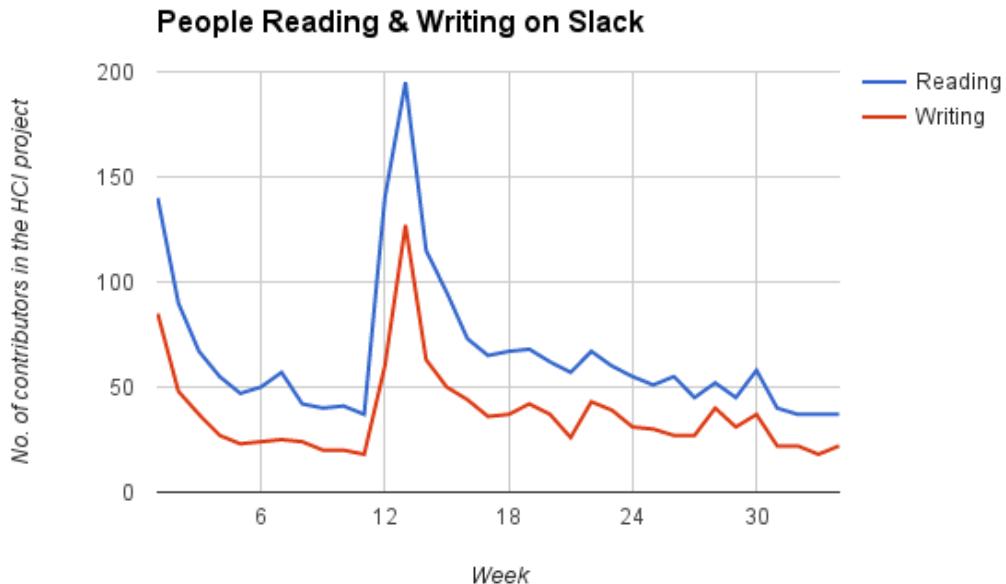


Figure 4.4: The measure of activity and participation in the HCI project: people reading and writing on Slack, collaborating towards discussions and paper writing. Slack was our primary channel of communication in the HCI project. Note the huge spike when HCI project recruited 373 more students in the second round.

4.4.1.2 Participation: Communication Activity - Piazza

Because the three projects were lead by three different professors, each had their own approach and preference of running the project. Instead of Slack, the Data Science and Computer Vision relied on Piazza for sharing instructions and notices.

In the first week, Computer Vision had 192 active participants, while Data Science had 127. Through their entirety, the activity for Computer Vision never fell below 15 viewers, while for Data Science it always stayed above 25. The low viewership was caused because participating students got busy with exams, the inability to follow

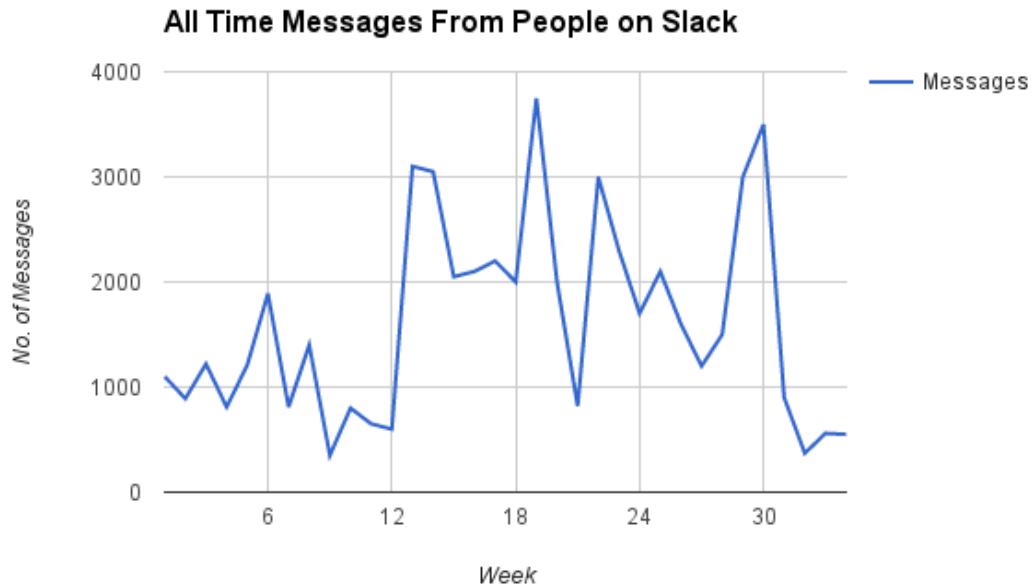


Figure 4.5: The measure of activity and participation in the HCI project: all time messages from people on Slack. Note high activity around second round recruitment (12th week), UIST submission (19th week) and CHI submission (30th week).

up afterwards also caused them to leave the project. By the end of 12 and 15 weeks respectively, Computer Vision still has 22.39% of the original active participants, while Data Science had 19.68% - see Figure 4.6.

4.4.1.3 Participation: Meeting Activity - Google Hangouts/YouTube

All three projects across this initiative relied on Google Hangout on Air for weekly meetings between advisor and crowd participants - that got recorded as YouTube videos by default for later viewing.

In total, the videos from three projects accounted for 8732 views, where 49%

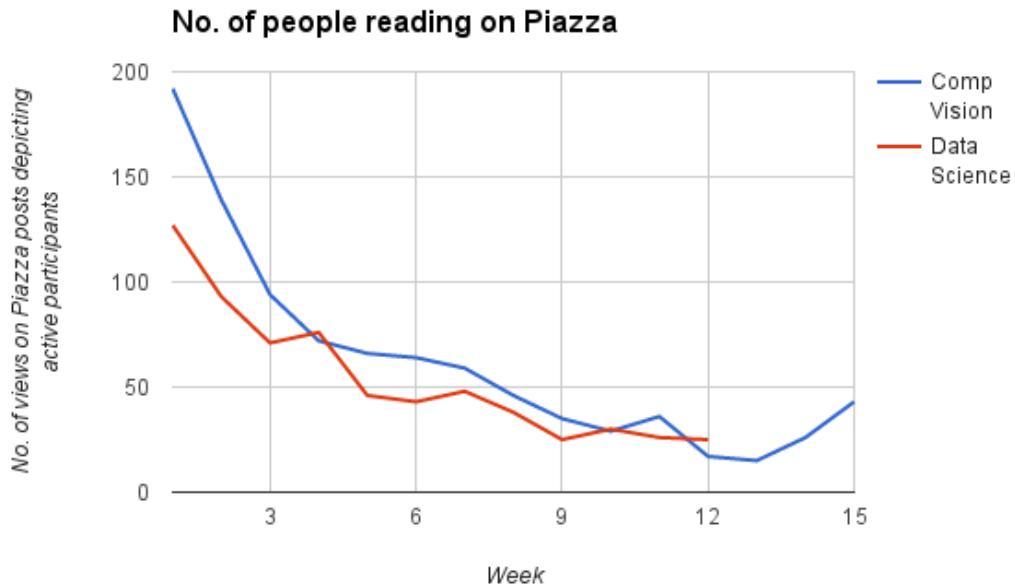


Figure 4.6: The measure of activity and participation in the Computer Vision and Data Science projects: number of people reading on Piazza. Data Science project ran for 12 weeks, while Computer Vision lasted for 15 weeks.

(4,304 views) came from India, while 30% (2,651 views) came from the US. On average, the HCI project that started three weeks after Comp Vision and Data Science - each of its video had 113.28 views, while median being 72 views within one week of hosting. As we can see in Figure 4.7, except for one, the viewership for any week hasn't gone below 25 views. For Computer Vision, the average and median views were 62.27 and 40.5 respectively, while for Data Science they were 65.5 and 43.5 respectively.

Another measure of activity is engaging with the weekly video conferencing meetings. In total, our participants viewed more than 125,000 minutes of video, where average view duration being 14 minutes and 18 seconds. In HCI, each video was watched

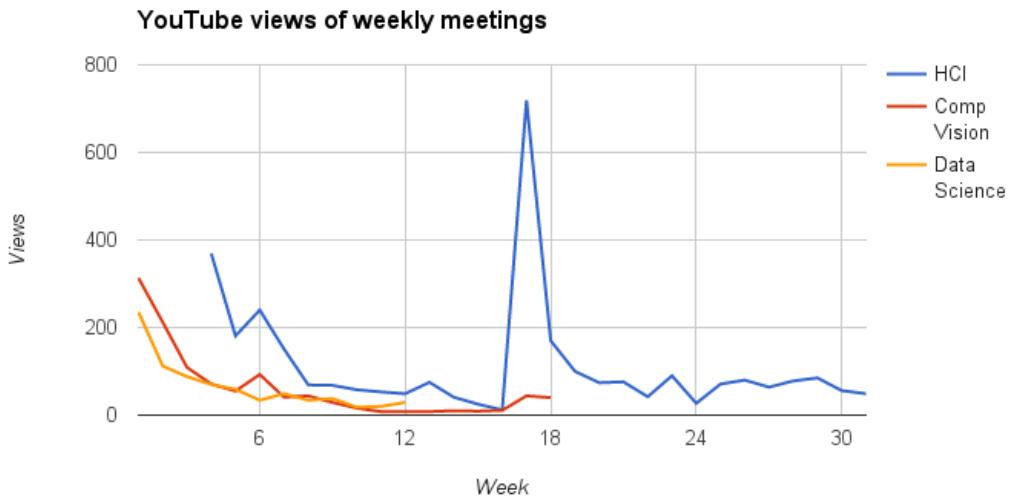


Figure 4.7: The measure of activity and participation across all projects: meeting views on YouTube from one of week of hosting. Note a huge spike for the HCI project, caused by the second recruitment.

for an average of 2,565 minutes (median of 1798.5 minutes), where except one, none of the videos were watched for less than 500 minutes (16th week). In Computer Vision, each video was watched for an average of 1,123 minutes (median of 459.5), while in Data Science, the average time was 1,221.08 minutes (median of 584.5 minutes) - see Figure 4.8. Inspired by its effectiveness, we also observed self-organized meetings by the crowd participants in the HCI group. During a span of about 30 weeks, we saw 41 videos produced by crowd organized meetings, with average views being 29.05.

4.4.1.4 Participation: Other Activities - Wiki and Github

In the HCI project, we used wiki as our primary medium for sharing instructions, where participants also added their milestone submissions. In the span of about

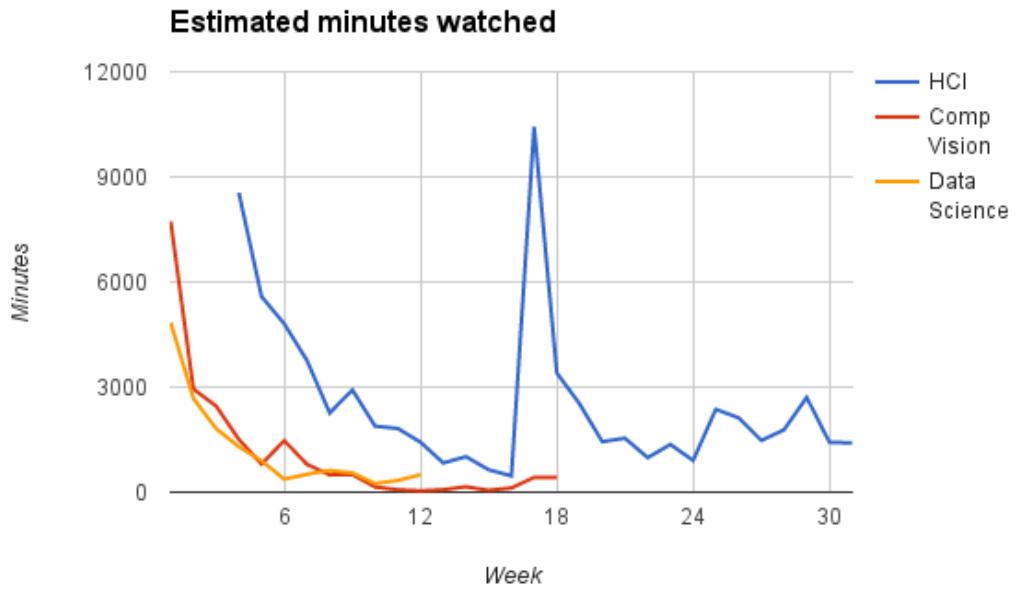


Figure 4.8: The measure of activity and participation across all projects: estimated minutes on YouTube from one of week of hosting. Note a huge spike for the HCI project, caused by the second recruitment. All the videos were unlisted, to prevent viewing by people outside our network. People who were unable to attend live meeting, viewed this video later.

30 weeks, 229 content pages have been published, with 12,128 edits and 895 files have been uploaded - mostly by crowd. These files range from prototype designs to research reports. During this time, the wiki has attracted more than 144, 000 views, with 11.88 views per edit. For our project, wiki has proven to be very critical in collaborative writing and milestone submissions.

HCI project also focused towards building the system [34] by the crowd. Over the span of 25 weeks, 1,564 code commits were made. As we can see from Figure 4.9, the number of commits increased to 135 a week after second round of recruitment (13th

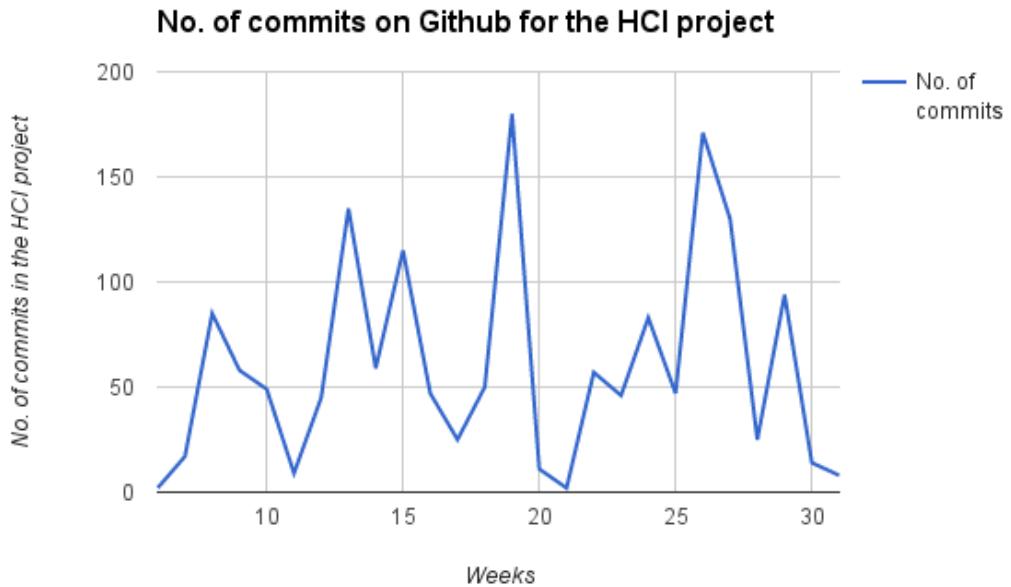


Figure 4.9: The measure of activity and participation in the HCI project: number of Github commits. Note the spikes around 13th (second round of recruitment), 19th (around UIST deadline) and 26th week (preparing for studies for CHI deadline).

week), and saw a bump during the UIST paper deadline (19th week). A major surge in commits also occurred in the 26th week, four weeks before CHI deadline, to ready the system for running studies and experiments. This pattern further reinstates that even though the research projects ran much longer than MOOCs, they continued to involve participants who got together enthusiastically around the deadlines.

4.4.1.5 Longevity: MOOC and The Aspiring Researchers Challenge (ARC)

For voluntary efforts such as MOOCs and citizen science, it is extremely challenging to keep participants motivated to continue to work. Other time commitments

(personal or work), lack of necessary background or skills, unfair credit distribution or other reasons can easily cause a participant to leave the campaign. In MOOCs, where courses usually last from 6 to 12 weeks, it's been reported that roughly 5-6% of students who signed up, earned a credential signifying official completion of the course [83, 110].

In contrast, our projects have lasted much longer with relatively lower attrition rate. Because of the nature of research projects, our criteria of contribution wasn't on the lines of courses where assignment submissions and final exams determined the activity. In our initiative, it ranged across brainstorming, discussions, analysis, prototyping, engineering and paper writing. This critical difference prevents us from comparing against MOOCs directly. Hence, we define attrition when a participant stops following the project or discussions, that leads to concrete contributions - like a paper.

In a MOOC, final exam is the culmination point; in research we believe authoring a paper is. Taking that into consideration, we found that - for HCI project, 9.25% of participants who signed up, co-authored a UIST paper in 19th week, while 8.42% of the participants co-authored a CHI paper in 30th week. For Computer Vision project, 18.91% of the participants co-authored an HCOMP paper in 15th week, while for Data Science, 32.40% of the participants were co-authors of another UIST paper in the 12th week. Table 4.1 states this data in the order of duration (weeks). Please note that, at the end of these specific weeks, not all authors were active, but their overall contribution to project was above a pre-set threshold - that made them eligible to be an author. This threshold was decided by the faculty and RAs, based on the PageRank score of credit distribution - to be discussed in detail in 'Credit Distribution and Team

Dynamics' section.

Projects and Initiatives	Duration	Participants through Completion
MOOC Final Exam	12 weeks	5%
ARC Data Sci.-UIST	12 weeks	32.4%
ARC Comp Vision-HCOMP	15 weeks	18.91%
ARC HCI-UIST	19 weeks	9.25%
ARC HCI-CHI	30 weeks	8.42%

Table 4.1: Percentage of people who completed or helped with the last milestones. Assuming that MOOCs take 12 weeks on average (it could be less), we can see that though our projects took longer for final completion (writing a paper vs. final exam) - the drop off rate was lower.

4.4.1.6 Longevity: Dropout Reasons

Overall, though our structure helped in producing papers and lowering drop off rate, we wanted to understand primary reasons of why it occurred anyhow.

We conducted a survey to understand whether our structure and approach caused people to leave, or were there other reasons. Aggregating responses from all the three projects - Comp Vision 51, Data Science 58 and HCI 64 responses; we found that time commitment and inability to catch up after exams were the primary reasons for crowd participants to leave. Given the fact that Computer Vision and Data Science projects had primarily undergraduate students, for them university exams were a high priority, and while the project continued to make progress, after exams, it was hard for many to follow up. HCI on the other hand was more diverse, consisting of working professionals, for whom work and personal life time commitment caused to leave the project. See Figure 4.10 for more details.

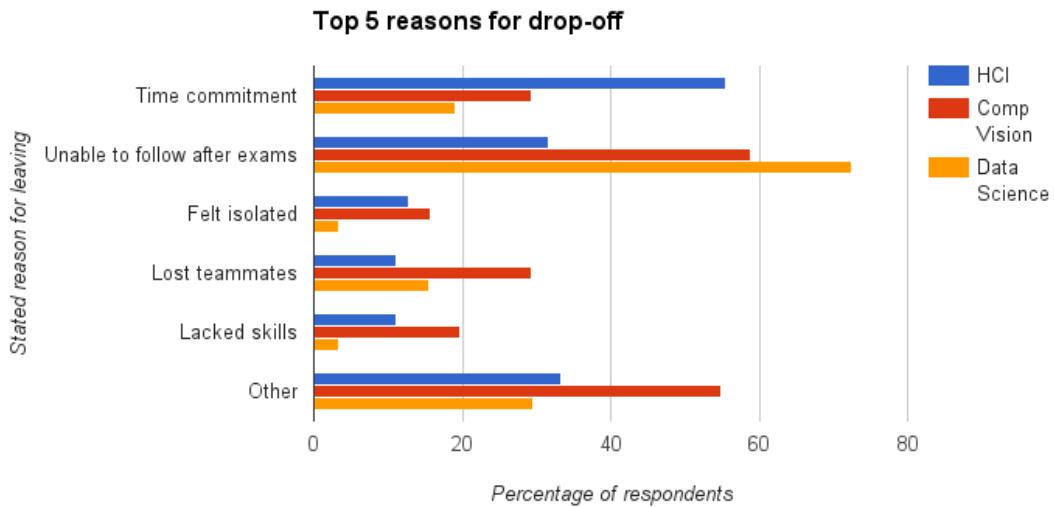


Figure 4.10: Top five reasons for people to drop-off in all three projects. Inability to follow the project after exams was the primary reason for Computer Vision and Data Science projects, while time commitment happens to be the primary reason for the HCI project. 'Other' reasons include: inability to understand the direction of the project, fatigue due to long duration, etc.

4.4.2 Paper Writing at Scale

Besides participating in brainstorming, coding and other aspects of the project, paper writing is an important research component. Despite the fact that about 80% of our participants had never co-authored a paper before, and had none to minimal research experience - we explored the possibility of crowdsourcing the paper writing process. Due to increased interest from participants in the HCI project, we let the crowd take the initiative. In Computer Vision and Data Science projects, papers were primarily written by research assistants (RAs), though crowd teams helped in providing reports of their individual work, that was later used and synthesized to write the paper.

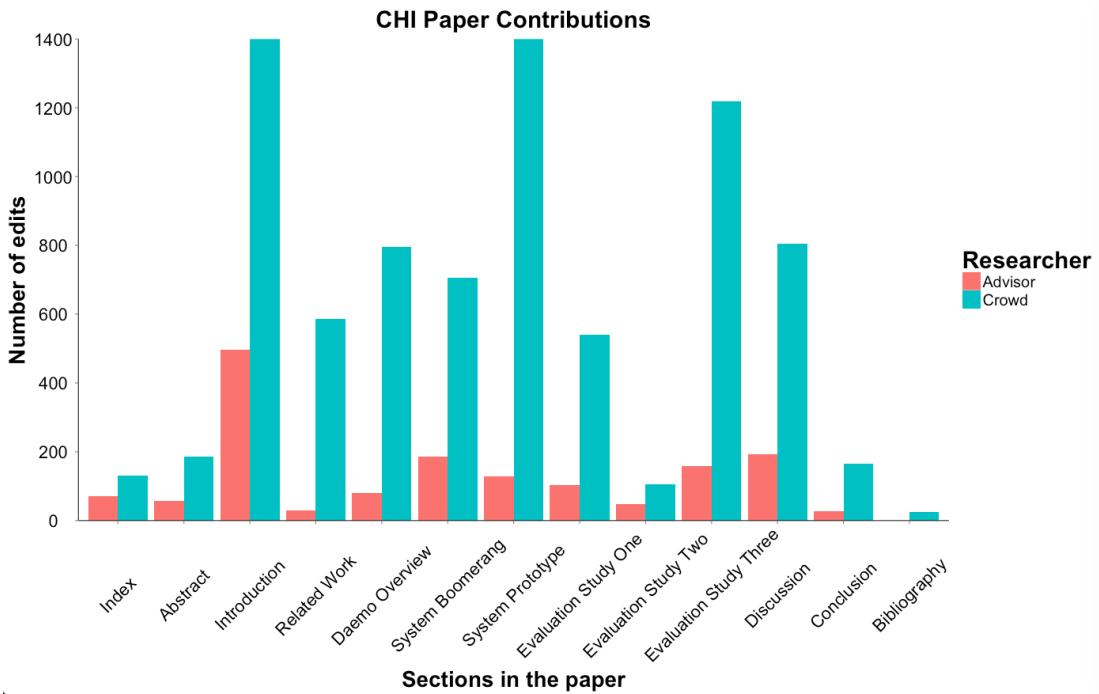


Figure 4.11: In the HCI project, crowd and faculty collaborated to write the CHI paper. This figure shows that crowd dominated the writing effort with about 4.5 times more edits than the advisor - a rare achievement, given the fact that more than 80% of the participants had never written a paper before.

4.4.2.1 The Paper Writing Process

To facilitate productive collaboration in the HCI project, we set up an online editable document and laid out the section level structure. This provided a starting point for the crowd to work on, where some edited each other's work, while some added their version next to each others using their identification id. These options were presented as guidelines for participants who felt awkward overwriting each other's work.

To reduce the dependency on advisor, we encouraged self-organized meetings among participants and assigned section wise DRIs. This formed a core group of writers

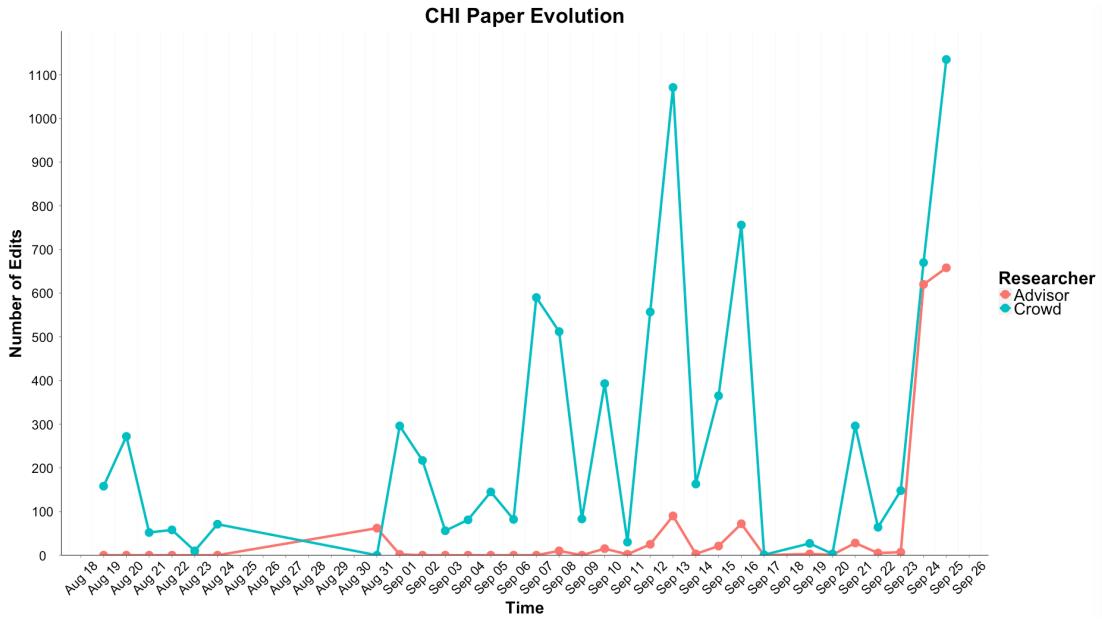


Figure 4.12: In the HCI project, crowd and faculty collaborated to write the CHI paper. This figure shows the activity of crowd vs. advisor throughout the duration of paper writing process. Though crowd did most of the work, the advisor became extremely active around the final deadline - helping synthesize the paper.

who worked with rest of the participants, and adopted a real-time editing approach while meeting over video conferencing. To synthesize the writing effort further, advisor was involved after every major writing session. Along with suggesting changes via comments, advisor also edited the paper directly. For the CHI paper, advisor made about 1,800 edits, while crowd made about 8,200 edits in about a month. Most of the edits by advisor were on the introduction section, which is a norm when advisor works with their graduate students too (see Figure 4.11). Pitching the story is a hard task, that requires expertise. In our paper writing experience, that's where crowd needed most help, though the total number of edits for the introduction section were still more than

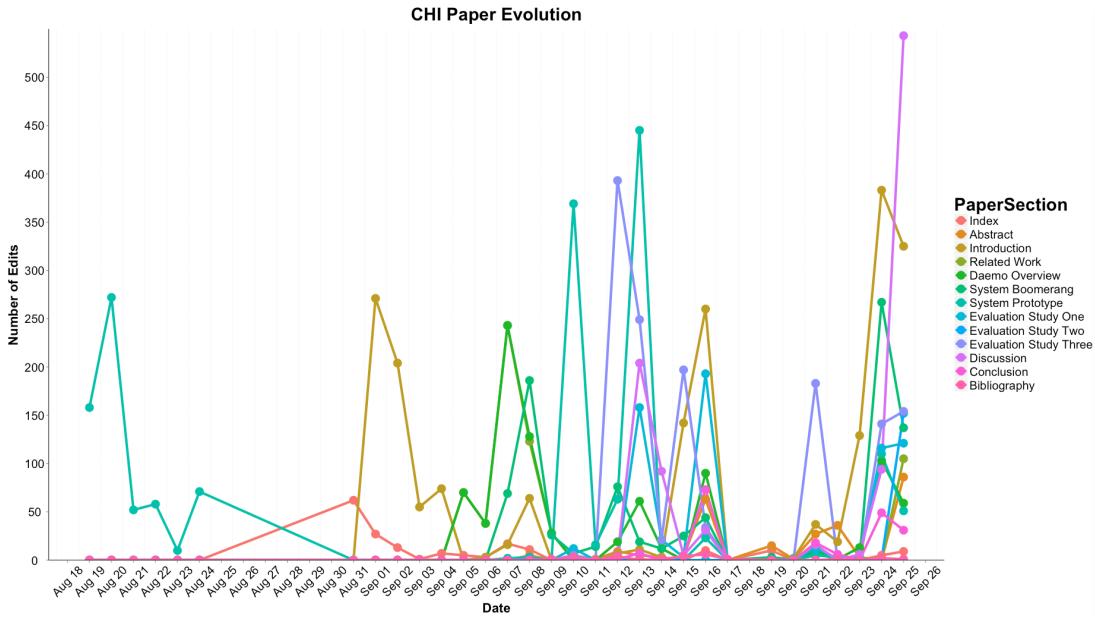


Figure 4.13: In the HCI project, crowd and faculty collaborated to write the CHI paper. This figure shows section wise evolution of the paper - starting with 'overview' and ending with 'discussion' - a very standard approach to write a paper.

that of advisor. For rest of the sections, advisor played an important role in synthesizing the direction of the paper, though they were primarily written by the crowd. Upon conducting a survey with 64 HCI participants, we found that 81.6% of them found the paper writing experience to be educational and enjoyable (median = 5). And 80.8% liked the involvement and role of DRIs during the paper writing process (median = 4).

4.4.2.2 The Paper Writing Evolution

The effort for writing the paper began almost a month from the deadline of 25th September 2015. The advisor had set an internal deadline, 10 days before actual deadline to create momentum and to give crowd sufficient time to refine upon the

writing. Though the crowd remained active throughout, we can see high activity of advisor in the initial setup phase, then around the internal deadline and then before the final deadline. As shown in Figure 4.12, this behavior is very typical to and reflects the collaboration between an advisor and their graduate students. In essence, it reflects the role of advisor in synthesizing the direction of the paper, while majority of it was written by the crowd.

Looking at the timeline for section wise progression in Figure 4.13, we rediscover high activity around internal and final deadlines - where the initial activity was for the overview, while the ending was for discussion.

4.4.3 Credit Distribution and Team Dynamics

For a paper with multiple co-authors, setting up author order can be challenging, yet an important research component. Incorrect or unfair order can demotivate participants. Therefore, unlike traditional approaches, we democratized the effort in the HCI project - we let crowd decide. However, as crowd was not directly involved in the paper writing effort for Computer Vision and Data Science projects, the credit distribution was simpler for them.

In the Computer Vision project, the reports submitted by the participants were synthesized by the RAs. Therefore, RAs became the earlier authors, followed by 19 top contributors (in alphabetical order), followed by 30 remaining contributors (in alphabetical order). In Data Science project, two participants made outstanding contributions and became top authors, followed by the RAs and advisor, followed by

rest of the participants in alphabetical order. In this section, we'll discuss the credit distribution mechanism in the HCI project. This form of credit distribution also unveiled multiple team dynamics within the community - this section walks through that.

In past, for large authored papers such as one by Tomlinson et. al [124] - the author sequence was ordered by placement "bids" from authors, with the first author's judgment used to resolve ties. In this past effort, the authors were experienced researchers and acquaintances, who merely contributed in writing, so it was easier for the first author to take the lead. In our initiative, the contributions were highly varied, and there were more than 50 author candidates - making it difficult for anyone to break the ties.

To implement our democratized approach in the HCI project, we tried two approaches with slight variation for two papers - at CHI and UIST respectively.

4.4.3.1 CHI - hidden distribution via a form, no faculty/RA involved in credit distribution

For CHI paper's author order, we set up a form and assigned 100 credits to all 60 author candidates who were active at the time. These authors were asked to distribute 100 credits among other participants on the basis of their knowledge of contribution. For example, if there were five participants in total, and one was asked to distribute the score among four, the participant might spread out the points as 76-12-12-0, if they thought one to be the lead contributor, two others contributed a little bit, and one made no contribution. Participants were free to give contributors no points if they did not

recognize them or did not know of their contributions. We collected this data over a span of two weeks and got 1,165 credit exchanges among 60 author candidates.

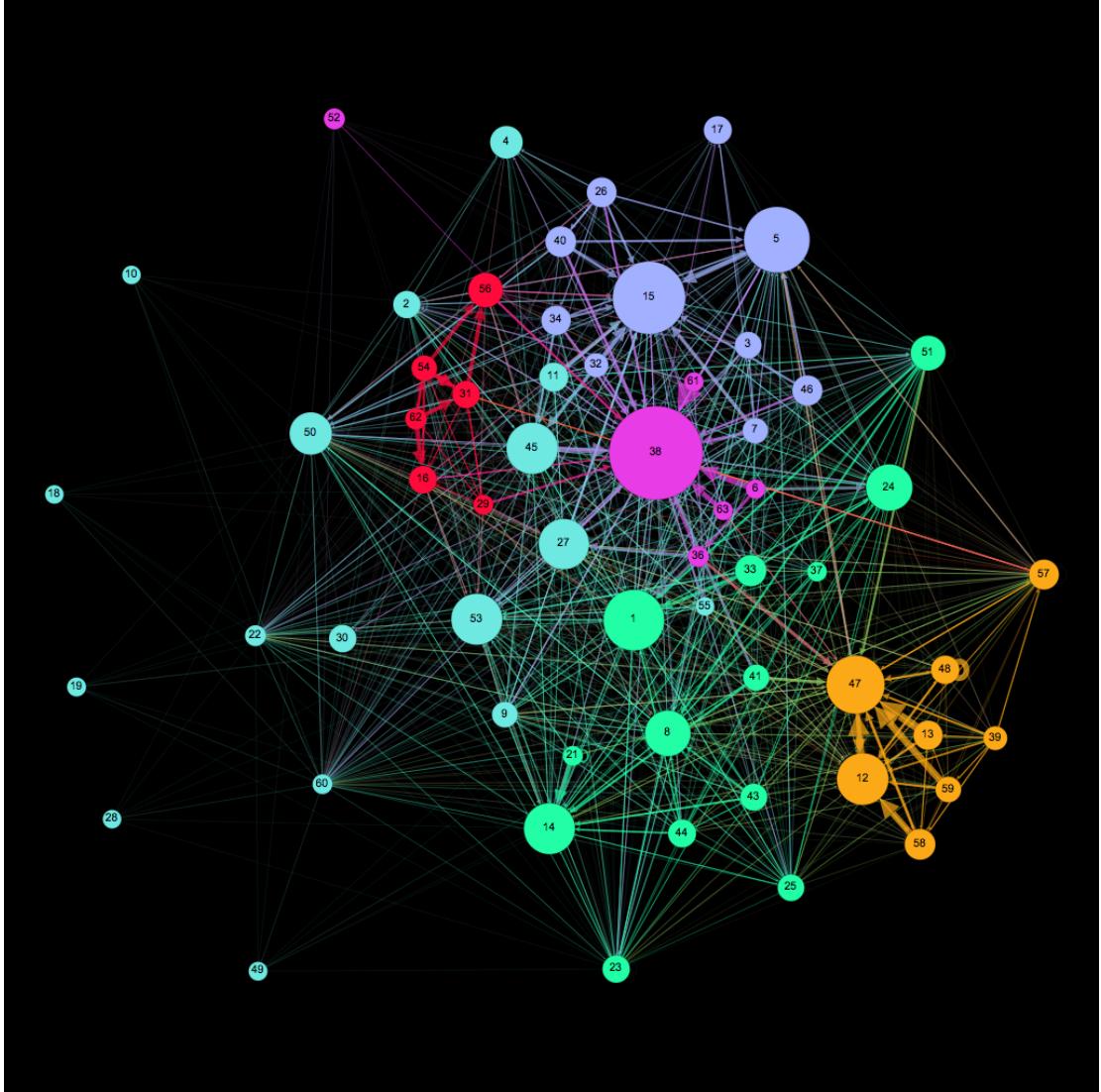


Figure 4.14: Credit distribution network graph in the HCI project for CHI paper: size of the nodes determine the PageRank, largest being the top author (node# = 38); width of the arrows determine the score sent from node A to B. Different color represent different groups, based on expertise/interest or cultural binding. Note higher scores exchanged between orange and red clusters among each other, while the largest node (node# = 38) receives scores of varying size from multiple sources.

Once we have everyone's credit points, we ran the PageRank [101] algorithm on top of the credit network to determine a credit ranking which we can use for authorship. PageRank works by finding individuals who everyone gives high credit to, then gives them more weight in the final credit assignment. So, even if small groups of people are all voting for each other to try and increase their credit, PageRank will realize that nobody else is giving them credit, so the aggregate impact of their credit assignment will be minor. This helps in addressing intra team credit biases. To allow only top contributors, we set a PageRank threshold where seven of the author candidates couldnt make it - totalling to an author list of 53 participants. The advisor and their RAs were listed last in the author order and added four more to the list of 53 authors.

To understand PageRank and other team dynamics, we can refer to Figure 4.14 for details. The size of nodes in the figure reflect the PageRank, while the width of edges determine the amount of score distributed. Through this image, we can derive following information:

- Based on expertise/interest, people form new groups. We can see new groups of technical people (blue, pink), design people (cyan, pink) and governance (green, pink) being formed - these groups tend to submit *inter* group votes.
- Groups in orange and red are formed out of cultural or personal connections, and tend to submit *intra* group votes. In such groups, individuals gave highest votes to each other. Note the weighted edges inter-between orange and red nodes.
- PageRank helps people who are contributing and driving multiple fronts (engi-

neering, design, governance, operations), and likely to get high ranks. To prove this further, we can notice the effect of PageRank through two clusters around nodes 38 (author ranked 1 in pink) and 47 (author ranked 5th in orange). We notice that though node 38 received multiple scores of all sizes, node 47 received fewer but higher scores - from within its group. This in effect was caused due to a group of participants working closely with each other and lead by node 47. According to pure aggregated scores, node 47 would rank second with 494.5 points, while node 38 had 669. Second, third and fourth authors scored 407.5, 251.5 and 258.25 respectively. However, after applying PageRank, the effect of intra team scores distribution diluted, and the scores given by node 38 played an important role. Thus, making score distribution democratic and representative of majority, not small clusters of participants.

Overall, we learn that an individual working across various disciplines tend to get diverse votes resulting in larger PageRank.

4.4.3.2 UIST - public distribution via badges, faculty/RA involved in credit distribution

Unlike in CHI, for UIST, we implemented public badge distribution system. As seen in Figure 4.15, the *intra* group credit exchange was significantly reduced (except when node 21 gave high score to node 6) than CHI; and more distributed scores were exchanged. To encourage badge distribution, RAs were also involved (see red), who distributed badges among all, for a variety of contributions - this created a culture of

acknowledging contributions via badges.



Figure 4.15: Credit distribution network graph in the HCI project for UIST paper: size of the nodes determine the PageRank, largest being the top author (node# = 26); width of the arrows determine the score sent from node A to B. Different color represent different groups, based on expertise/interest or cultural binding. The largest node in red is an RA, who distributed badges to all and create an ecosystem for acknowledging via badges. Only one example of major score exchange is observed in the intra group in orange - significant reduction from CHI, possibly due to the public nature of distribution.

To understand both the approaches better, we conducted a survey with HCI participants, and received 64 responses. On a Likert scale of 5, we learned that though participants were satisfied with their author order in both CHI and UIST (median = 4); they liked distributing credits via public badges (median = 4) over hidden form (median = 3). This is also reflected when comparing the nature of distribution in Figure 4.15 and Figure 4.14.

4.5 Survey Findings and Discussion

The Aspiring Researchers Challenge was inspired by primarily two research questions:

- Can we motivate people around the world to come together and do scientific research with experts?
- Can we crowdsource an open ended problem and provide access and educational value at scale?

Over the span of about six months, the initiative has motivated more than 1,000 people from around the world to do research in different areas of computer science. Of them, more than 150 people have contributed towards co-authoring papers at top-tier conferences. The program essentially emulated a research lab at scale, and it is valuable to understand and recommend a variety of its aspects. As an effort, we conducted a survey across all the three projects. In total, we got 51 responses in Computer Vision project, 58 in Data Science and 64 in HCI.

The prime motivation for people to join this initiative was to gain research experience (Computer Vision = 92.2%, Data Science = 98.3%, HCI = 85.9%). After many weeks into each project, we found that people enjoyed and learned about research, academia, and research topics in general. On a Likert scale of 5, people responded with a median of 5 for the HCI project, while 4 for Computer Vision and Data Science. Providing scaffolding and educational value to these aspiring researchers was one of our prime objective, and the initiative seems to have provided and met their motivations.

To make this possible, we addressed participants as researchers and not as smart workers on crowd market places. We encouraged them to read academic papers, taught research topics via video meetings, trained them in required skills and attempted to match the overall process to that of a graduate school and bring similar experience.

More than 80% of the participants had never co-authored a paper before. However, our structure and program design provided various contribution opportunities. Participants in the HCI project found contributing to writing the UIST/CHI papers to be an enjoyable and educational experience (median = 5). They believe that participating in this program has increased their interest in pursuing research in the future (median = 5) and that this program has provided valuable experience and strengthened their resume (median = 5).

For any research project, communication and collaboration are key to its success. In this initiative, we facilitated it through multiple mediums as discussed above - Slack, Google Hangout on Air, Wiki and Piazza. Though different projects used platforms of their choice, better communication fostered collaboration. We encouraged

self-organized meetings, constructive peer-feedback, acknowledgement badges as means to build a positive ecosystem. Instead of one way communication as in MOOC, we encouraged a culture where faculty would guide rather than manage, and engage in conversation to echo and synthesize critical discussions.

Upon asking participants in the HCI project, we learned that our program facilitated lot of communication and brainstorming opportunities (median = 5), and that they enjoyed direct video discussions with the faculty (median = 5). Video meetings created a sense of personal connection, and emulated traditional advisor-student relationship. Participants also found faculty and RAs to be very helpful, approachable and understanding (HCI median = 5, Computer Vision median = 4, Data Science median = 4). This reflects from the fact, that when we encouraged goal oriented ad-hoc teams in the HCI project - crowd enjoyed forming teams and working with other participants (median = 4). They liked coordination and collaboration between different teams working on coding, research and user-studies (median = 4); and rallied together during the paper deadline periods (median = 4).

For crowd to be able to contribute effectively, the project design plays a very critical role. A directionless project can confuse the participants, while a rigid project can cause people to dropout. We believe that a fault tolerant project, that accommodates people with varying skills and allow them to resume when possible - proves to be more successful. The design takes into consideration, that people have their jobs and exams, for which they might be unavailable sometimes. In our initiative, the HCI project is extremely fault-tolerant, where multiple people work in parallel or collaborate

to make progress - hence, also ongoing for more than 30 weeks.

Based on survey input and experiences, we learned the importance of fault tolerant design of projects. We learned that crowd should be given freedom to act and collaborate, the more they collaborate, the higher PageRank they'll have - making them top author. We also believe that interactive communication and meeting with personalized touch can cause great educational value and motivation for crowd.

4.6 Conclusion and Future Work

This chapter presents the first of its kind research-at-scale initiative - The Aspiring Researchers Challenge. As part of this initiative, we attempted to solve open ended research problems by crowdsourcing; and providing access and educational value to people around the world. Over the span of six months, the initiative attracted more than 1,000 participants from 6 continents, to work on projects in computer science - computer vision, data science and human-computer interaction (HCI). These projects were lead by faculty at Stanford and UC Santa Cruz, and by following our structure, lead to three top tier work-in-progress papers at ACM UIST and AAAI HCOMP 2015. A full paper by HCI participants is under review at ACM CHI, while two graduate school aspirants earned an RA position at Stanford. Based on our survey findings, we learn that participants found this endeavor to be educationally valuable, and the program has increased their interest in pursuing research in the future.

The initiative is still ongoing with HCI project being active. We plan to make

another round of open call and release the alpha version of the system (Daemo [34]) - making real world impact. With changing dynamics of having Daemo used by people around the world, crowd coordination and Daemo management will change - causing us to explore and design new structure and organizational process. We believe that this initiative is a first step towards realizing research on a wider scale - and we can invite more open ended problems to be solved, while providing educational value.

Appendix A

The Aspiring Researchers Challenge: Previous Work

In chapter 4, I reflected motivations, experiences and findings of the Aspiring Researchers Challenge [38]. The initiative in its current form was built upon experiences from two pilot studies - both run in a class set up at UC Santa Cruz. In the first pilot study, we produced a published paper [116] at an IEEE conference [15] - entirely written by students. In the second pilot study, we produced better results for an open ended problem being addressed by Stanford Vision Group. This appendix contains a brief overview of the two pilot studies, that helped build an understanding and lead to a larger initiative.

A.1 Pilot Study 1

Building upon a previously proposed framework [125], we conducted our first pilot experiment, aiming to produce a crowdsourced publishable paper. The controlled experiment was conducted in a computer graphics class for graduate students in the Winter quarter at UC Santa Cruz in 2014. In the beginning of the class, the professor proposed three research problems to work on, and the class was divided into three major groups. The process involved students selecting the research problem, working on the implementation, generating results and writing sections of the paper in parallel to encourage diversity [64]. The students contributed by writing sections in the order of Related Work, Implementation, Methods and Results and finally Introduction and Conclusion. By applying iterative peer-grading techniques, the best sections and results from each student were selected and made it to the paper. Interactive periodic meetings with the professor were conducted in the class, regarding the direction of the project and next steps. Unlike individual or direct feedback meetings between PI and students, these meetings were on a broadcast level - interacted only at a top level with the peer-graded ideas.

This process resulted in three papers with complete content but unstructured flow from section to section. On the basis of student engagement and interest, the paper with maximum activity was selected to be worked upon further. To optimize the flow of the selected paper, a paper-a-thon was conducted post quarter, bringing together all interested students. After further refinement, the paper was pre-peer reviewed twice by a

group of four senior graduate students before sending to a relevant IEEE conference [15].

The paper got accepted [116].

During this process, we learned that handling a project from idea inception to paper writing is a non-trivial task. However, post paper-a-thon efforts proved to be extremely useful. At the end of this endeavor, we achieved the goal of getting a paper accepted, written by minimally supervised students. We learned about student motivations, team dynamics and coordination approaches that made it possible. Encouraged by the overall success, and as part of the eventual goal of crowdsourcing research, we next focused on collaborative crowd research for producing publishable results.

A.2 Pilot Study 2

Based on the learning from pilot study 1, we conducted second round of controlled experiment in an undergraduate class in Spring quarter at UC Santa Cruz. To implement our learning, we focused on the research process part instead of paper writing; and recruited students interested in research.

In collaboration with the Stanford Computer Vision Group, we proposed an open-ended research problem to the selected group of undergraduate students at UCSC. The project is in progress and run by graduate students at Stanford, with few versions live. Undergrad students were blindfolded from the approaches used by graduate students, and encouraged to come up with their own ideas. These ideas would later be compared against methods developed at Stanford by graduate students. The process in-

volved team formation, where four teams were formed to work towards a common problem using their unique approach. Teams worked through a series of research phases: brainstorming, paper-pencil prototyping, development and user-evaluation. To boost productivity across these stages, student crowd and process were managed by Stanford and UC Santa Cruz researchers using an earlier proposed framework rules [125]. Using peer-grading techniques, the best approach per-phase determined researchers feedback to all the participating teams. Therefore, also encouraging others to iteratively improve their contributions. Interactive periodic meeting was conducted online over a video conferencing software.

By the end of the quarter, the solution proposed by the student crowd with no computer vision background compared reasonably well against the ones developed by a senior PhD students with expertise in the area. To reach this conclusion, undergraduate students were taught qualitative and quantitative analysis methods - who ended up conducting the experiments and analysis. We learned that the framework proposed to manage student crowd and research process is effective. Moving forward we planned to reach out to larger crowd with bigger goals [38].

Appendix B

The Aspiring Researchers Challenge: Crowd Publications

In chapter 4 and Appendix A, I mentioned about the projects done by crowd that got accepted as full and work-in-progress papers. However, due to emphasis on the meta level aspects of the initiative, those were not discussed. To understand the scope and scale of the projects, I'll provide a brief overview of the crowd work that got published.

B.1 Daemo: a Self-Governed Crowdsourcing Marketplace

B.1.1 Meta Information

This project is part of the human-computer interaction (HCI) research under the Aspiring Researchers Challenge - lead by Stanford professor Michael Bernstein. 61 participants helped co-author a work-in-progress paper [70] that got accepted at ACM

Symposium on User Interface Software and Technology (UIST 2015). Later, 53 of those participants helped co-author a full paper for ACM Conference on Human Factors in Computing Systems (CHI 2016) - under review.

B.1.2 Overview

Crowdsourcing marketplaces provide opportunities for autonomous and collaborative professional work as well as social engagement. However, in these marketplaces, workers feel disrespected due to unreasonable rejections and low payments, whereas requesters do not trust the results they receive. The lack of trust and uneven distribution of power among workers and requesters have raised serious concerns about sustainability of these marketplaces. To address the challenges of trust and power, this project introduces Daemo, a self-governed crowdsourcing marketplace. We proposed a *proto-type task* to improve the work quality and *open-governance model* to achieve equitable representation. We envisage Daemo will enable workers to build sustainable careers and provide requesters with timely, quality labor for their businesses. The system's website is live at daemo.stanford.edu.

B.2 On Optimizing Human-Machine Task Assignments

B.2.1 Meta Information

This project is part of the computer vision research under the Aspiring Researchers Challenge - lead by UCSC professor James Davis and Cornell professor Serge

Belongie. 49 participants helped co-author a work-in-progress paper [128] that got accepted at AAAI Conference on Human Computation & Crowdsourcing (HCOMP 2015).

B.2.2 Overview

When crowdsourcing systems are used in combination with machine inference systems in the real world, they benefit the most when the machine system is deeply integrated with the crowd workers. However, if researchers wish to integrate the crowd with “off-the-shelf” machine classifiers, this deep integration is not always possible. This work explored two strategies to increase accuracy and decrease cost under this setting. First, we showed that reordering tasks presented to the human can create a significant accuracy improvement. Further, we showed that greedily choosing parameters to maximize *machine* accuracy is sub-optimal, and joint optimization of the *combined* system improves performance. We found that Task Ordering can have a dramatic impact on the human time budget required to achieve a particular target accuracy. A cost savings of 62% was achieved in our experiments.

B.3 Investigating the ”Wisdom of Crowds” at Scale

B.3.1 Meta Information

This project is part of the data science research under the Aspiring Researchers Challenge - lead by Stanford professor Sharad Goel. 58 participants helped co-author a work-in-progress paper [117] that got accepted at ACM Symposium on User Interface

Software and Technology (UIST 2015).

B.3.2 Overview

In a variety of problem domains, it has been observed that the aggregate opinions of groups are often more accurate than those of the constituent individuals, a phenomenon that has been termed the “wisdom of the crowd”. Yet, perhaps surprisingly, there is still little consensus on how generally the phenomenon holds, how best to aggregate crowd judgements, and how social influence affects estimates. We investigated these questions by taking a meta wisdom of crowds approach. With a distributed team of over 100 student researchers across 17 institutions in the United States and India, we developed a large-scale online experiment to systematically study the wisdom of crowds effect for 1,000 different tasks in 50 subject domains. These tasks involved various types of knowledge (e.g., explicit knowledge, tacit knowledge, and prediction), question formats (e.g., multiple choice and point estimation), and inputs (e.g., text, audio, and video). To examine the effect of social influence, participants are randomly assigned to one of three different experiment conditions in which they see varying degrees of information on the responses of others. In this ongoing project, we are now preparing to recruit participants via Amazon’s Mechanical Turk. The system’s website is live at wisdomofcrowds.stanford.edu.

B.4 RTI Compression for Mobile Devices

B.4.1 Meta Information

Part of pilot study 2, this project was conducted in a class set up at UCSC - lead by professor James Davis. More than 10 people made contributions to the project. However, four most active co-authored it together with the staff. The paper [116] got accepted at IEEE International Conference on Information Technology and Multimedia (ICIMu 2014) [15].

B.4.2 Overview

Currently, Reflectance Transformation Imaging (RTI) technology is restricted to desktop and high performance computing devices. In recent years, mobile devices and tablets have become ubiquitous due to their increased performance. However, the size of RTI files (≈ 100 MB) limits the range of portable devices capable of displaying RTI files. In this project, we explored different compression techniques, and developed an RTI viewer prototype for both Android and iOS devices. Experiments with JPEG, JPEG2000, lossless compression show the resulting error for compression ratios of 30:1 is comparable to the error of traditional two-dimensional images. For higher compression rates, we presented a novel $\alpha\beta$ -JPEG algorithm which compresses color and reflectance information individually.

Appendix C

The Aspiring Researchers Challenge: HCI Project - Idea Flow

In chapter 4, I discussed results from the Aspiring Researchers Challenge. While most of the results in all the projects were addressed, an important aspect of the HCI project - the idea flow is described here in detail.

As part of the coordination and idea generation strategy, we let crowd participants produce ideas in parallel every week. Followed by peer review and faculty-RA discussions, top ideas were considered and echoed back to the crowd during weekly meetings. Based on the ideas that were echoed, milestones were designed by the faculty and RAs, encouraging crowd participants to build further upon them. This practise was highly active and followed in the HCI project.

In the HCI project, it took eight weeks and five iterations to come up with four foundation ideas that lead the direction of the project. As seen in Figure C.1, in the first

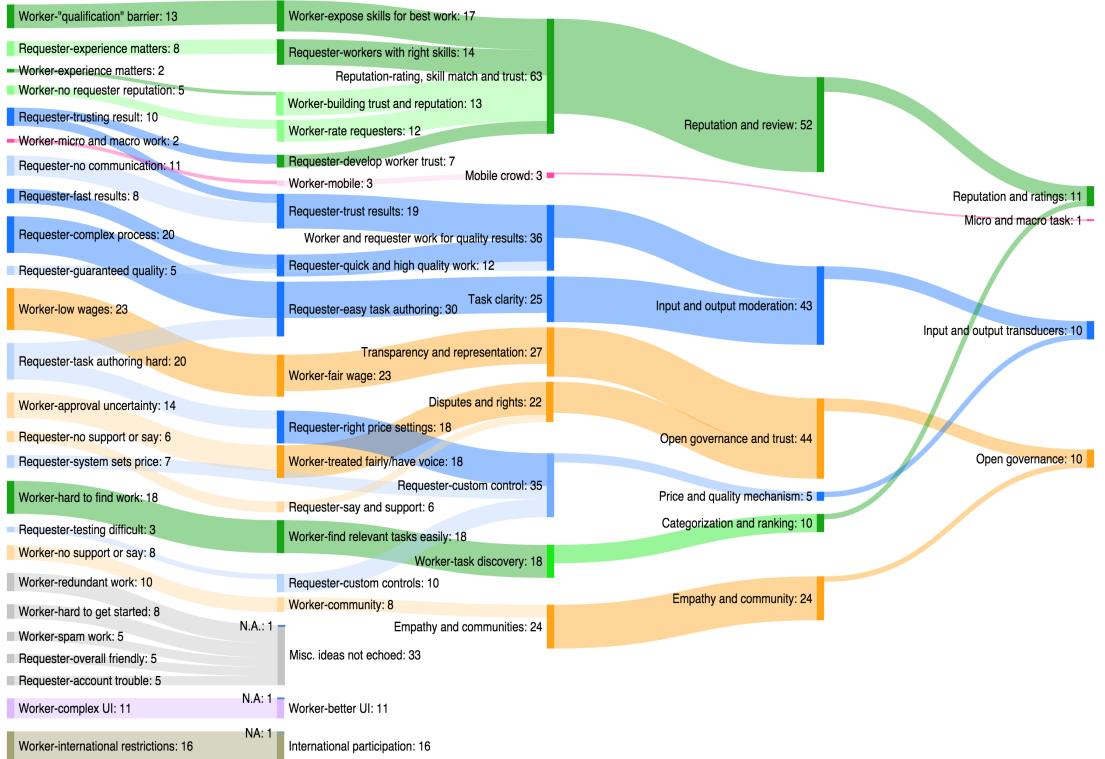


Figure C.1: Flow of ideas in the HCI project, from 23 categories to 4, in 5 iterations (8 weeks). The project was to build a new crowd market, that gives more power and features to workers and requesters. Darker shades represent ideas that faculty echoed in weekly meetings, while lighter shades represent ideas that were not reflected or echoed. Number represents, number of times a particular idea was proposed. Some ideas lasted a few iterations and did not contribute to the final four (bottom three categories). Though “Micro and macro task” appears to be less popular (proposed fewer times), it consistently ranked higher (people voted it highly) - hence, it made it to the final four.

iteration, there were 23 categories of idea submission, of which 10 were echoed (darker shade of color). During this week, crowd participants were asked to reflect on problems they encountered while using crowd platforms as a worker and as a requester. In the second iteration, there were 20 ideas, again 10 were echoed. During second iteration, crowd participants were asked what features or aspects they would like to have in a crowd platform. In the third iteration, nine categories of ideas were proposed, and six were echoed. In the fourth iteration, all seven ideas were echoed. Finally, in the fifth iteration, all four categories of ideas were echoed. Some ideas did not contribute to the final four (bottom three in the figure), while some were not proposed highly, but voted highly (Micro and macro tasks) - ranked number one, twice.

In Figure C.1, we can observe that multiple categories were related to similar theme (same color) - some got echoed (darker shade), some did not (lighter shade), but they eventually got merged with each other. In essence, the idea flow diagram reflects the process of ideas getting produced and merged, sometimes ignored - to reach a synthesized set of foundations that helps projects to succeed. It also reaffirms the nature of crowd contribution and benefits of parallel efforts.

Bibliography

- [1] Amazon Mechanical Turk. <https://www.mturk.com/mturk/welcome>. Accessed: 2015-11-18.
- [2] Amazon Mechanical Turk Worker Statistics. <https://requester.mturk.com/tour>. Accessed: 2015-11-18.
- [3] Bombings Trip Up Reddit in Its Turn in Spotlight. <http://www.nytimes.com/2013/04/29/business/media/bombings-trip-up-reddit-in-its-turn-in-spotlight.html>. Accessed: 2015-11-15.
- [4] Census of India - Mode of Communication: 2001-2011. <http://goo.gl/lUTNad>. Accessed: 2015-11-15.
- [5] CNN: Crowdsourcing volunteers comb satellite photos for Malaysia Airlines jet. <http://www.cnn.com/2014/03/11/us/malaysia-airlines-plane-crowdsourcing-search/>. Accessed: 2015-11-15.
- [6] Crowdsourced videos, photos could aid Boston blast investigations. <http://www.cnet.com/news/crowdsourced-videos-photos-could-aid-boston-blast-investigations/>

- blast-investigations/. Accessed: 2015-11-15.
- [7] DARPA Network Challenge. <http://archive.darpa.mil/networkchallenge>. Accessed: 2015-11-15.
- [8] dbpedia. <http://dbpedia.org/>. Accessed: 2015-11-15.
- [9] Freebase. <http://www.freebase.com/>. Accessed: 2015-11-15.
- [10] From Computing Research to Surprising Inventions. <http://research.microsoft.com/apps/video/default.aspx?id=183934>. Accessed: 2015-11-15.
- [11] Github. <https://github.com/>. Accessed: 2015-11-21.
- [12] Google Hangout on Air/YouTube. <http://www.google.com/learnmore/hangouts/onair.html>. Accessed: 2015-11-21.
- [13] Google Knowledge Graph. <http://goo.gl/6cCKVP>. Accessed: 2015-11-15.
- [14] GupShup Re-launches GupShup Messenger, Innovates for India. <http://www.siliconindia.com/news/technology/GupShup-Relaunches-GupShup-Messenger-Innovates-forIndia-nid-140878-cid-2.html/1>. Accessed: 2015-11-15.
- [15] IEEE International Conference on Information Technology and Multimedia at UNITEN (ICIMu 2014). <http://icimu.uniten.edu.my/index.php/about>. Accessed: 2015-02-02.

- [16] Indian Railways. http://en.wikipedia.org/w/index.php?title=Indian_Railways&oldid=657313685. Accessed : 2015 – 11 – 15.
- [17] Internet Live Stats. <http://www.internetlivestats.com/internet-users/>. Accessed: 2015-11-15.
- [18] Is Massive Open Online Research the Next Frontier for Education? <http://ucsdnews.ucsd.edu/pressrelease/ismassiveopenonlineresearchthenextforeducation>. Accessed : 2015 – 11 – 15.
- [19] iStockPhoto. <http://www.istockphoto.com/>. Accessed: 2015-11-15.
- [20] Kitten War. <http://www.kittenwar.com/>. Accessed: 2015-11-15.
- [21] Mechanical Turk: Now with 40.92% spam. <http://www.behind-the-enemy-lines.com/2010/12/mechanical-turk-now-with-4092-spam.html>. Accessed: 2015-11-18.
- [22] Microsoft Bing Satori. <http://binged.it/10nFdH0>. Accessed: 2015-11-15.
- [23] Microsoft's social "Whodunit" competition to begin in India. <https://in.news.yahoo.com/microsofts-social-whodunit-competition-begin-india-045205972.html>. Accessed: 2015-11-15.
- [24] Mobile telephone numbering in India. http://en.wikipedia.org/w/index.php?title=Mobile_telephone_numbering_in_India&oldid=657910549. Accessed : 2015 – 11 – 15.

- [25] MOOC Wikipedia article. https://en.wikipedia.org/wiki/Massive_open_online_course. Accessed : 2015 – 11 – 21.
- [26] MyHeartMap Challenge. <http://www.med.upenn.edu/myheartmap/>. Accessed: 2015-11-15.
- [27] Physics paper sets record with more than 5,000 authors. <http://www.nature.com/news/physics-paper-sets-record-with-more-than-5-000-authors-1.17567>. Accessed: 2015-11-15.
- [28] Piazza. <https://piazza.com/>. Accessed: 2015-11-21.
- [29] Police ”Facewatch” app targets London riot suspects. <http://www.independent.co.uk/news/uk/crime/police-facewatch-app-targets-london-riot-suspects- 7887778.html>. Accessed: 2015-11-15.
- [30] Poverty headcount ratio at \$2 a day (PPP) (% of population). <http://data.worldbank.org/indicator/SI.POV.2DAY>. Accessed: 2015-11-15.
- [31] Registry of Standard Biological Parts. http://parts.igem.org/Main_Page. Accessed : 2015 – 11 – 15.
- [32] Slack. <https://slack.com/>. Accessed: 2015-11-21.
- [33] Social whodunnit competition launches in India - tech - 31 January 2013 - New Scientist. <http://www.newscientist.com/article/mg21729025.500-social-whodunnit-competitionlaunches-in-india.html>. Accessed: 2015-11-15.

- [34] Stanford Daemo. daemo.stanford.edu. Accessed: 2015-11-21.
- [35] Telecommunications in India. [http://en.wikipedia.org/w/index.php?title=Telecommunications_n&oldid=656139185](http://en.wikipedia.org/w/index.php?title=Telecommunications_in_India&oldid=656139185). Accessed : 2015 – 11 – 15.
- [36] The Tag Challenge. <http://www.tag-challenge.com/about-the-contest/>. Accessed: 2015-11-15.
- [37] The Whodunit? Challenge by Microsoft Research India. <http://www.whodunitchallenge.com/>. Accessed: 2015-11-15.
- [38] UCSC+Stanford: The Aspiring Researcher Challenge. <https://aspiringresearchers.soe.ucsc.edu/>. Accessed: 2015-11-18.
- [39] Wikimedia. <https://www.wikimedia.org/>. Accessed: 2015-11-21.
- [40] Wikipedia. <https://www.wikipedia.org/>. Accessed: 2015-11-18.
- [41] Wired Magazine's Vanish Challenge. http://www.wired.com/vanish/2009/11/ff_vanish2. Accessed : 2015 – 11 – 15.
- [42] Young men, please sue the New York Post. <http://www.washingtonpost.com/blogs/erik-wemple/wp/2013/04/22/young-men-please-sue-the-new-yorkpost/>. Accessed: 2015-11-15.
- [43] Akseli Anttila and Younghee Jung. Discovering design drivers for mobile media solutions. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '06, pages 219–224, New York, NY, USA, 2006. ACM.

- [44] Eric P.S. Baumer, June Ahn, Mei Bie, Elizabeth M. Bonsignore, Ahmet Börütecene, Oğuz Turan Buruk, Tamara Clegg, Allison Druin, Florian Echtler, Dan Gruen, Mona Leigh Guha, Chelsea Hordatt, Antonio Krüger, Shachar Maidenberg, Meethu Malu, Brenna McNally, Michael Muller, Leyla Norooz, Juliet Norton, Oguzhan Ozcan, Donald J. Patterson, Andreas Riener, Steven I. Ross, Karen Rust, Johannes Schöning, M. Six Silberman, Bill Tomlinson, and Jason Yip. Chi 2039: Speculative research visions. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 761–770, New York, NY, USA, 2014. ACM.
- [45] Gerard Beenens, Kimberly Ling, Xiaoqing Wang, Klarissa Chang, Dan Frankowski, Paul Resnick, and Robert E. Kraut. Using social psychology to motivate contributions to online communities. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, pages 212–221, New York, NY, USA, 2004. ACM.
- [46] Yochai Benkler. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, 2006.
- [47] Frank Bentley and Edward Barrett. *A time to glance: Studying the use of mobile ambient information*. The MIT Press, 2012.
- [48] Michael S. Bernstein, Joel Brandt, Robert C. Miller, and David R. Karger. Crowds in two seconds: Enabling realtime crowd-powered interfaces. In *Proceedings of the*

24th Annual ACM Symposium on User Interface Software and Technology, UIST '11, pages 33–42, New York, NY, USA, 2011. ACM.

- [49] Michael S. Bernstein, Greg Little, Robert C. Miller, Björn Hartmann, Mark S. Ackerman, David R. Karger, David Crowell, and Katrina Panovich. Soylent: A word processor with a crowd inside. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 313–322, New York, NY, USA, 2010. ACM.
- [50] Michael S. Bernstein, Jaime Teevan, Susan Dumais, Daniel Liebling, and Eric Horvitz. Direct answers for search queries in the long tail. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 237–246, New York, NY, USA, 2012. ACM.
- [51] Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and Tom Yeh. Vizwiz: Nearly real-time answers to visual questions. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 333–342, New York, NY, USA, 2010. ACM.
- [52] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD '08, pages 1247–1250, New York, NY, USA, 2008. ACM.

- [53] Andrew T. Campbell, Shane B. Eisenman, Nicholas D. Lane, Emiliano Miluzzo, and Ronald A. Peterson. People-centric urban sensing. In *Proceedings of the 2Nd Annual International Workshop on Wireless Internet*, WICON '06, New York, NY, USA, 2006. ACM.
- [54] Manuel Cebrian and Andrea Vattani. Finding red balloons with split contracts: robustness to individuals selfishness. In *In ACM Symposium on Theory of Computing*, pages 775–788, 2012.
- [55] Kuang Chen, Akshay Kannan, Yoriyasu Yano, Joseph M. Hellerstein, and Tapan S. Parikh. Shreddr: Pipelined paper digitization for low-resource organizations. In *Proceedings of the 2Nd ACM Symposium on Computing for Development*, ACM DEV '12, pages 3:1–3:10, New York, NY, USA, 2012. ACM.
- [56] Justin Cheng, Jaime Teevan, Shamsi T Iqbal, and Michael S Bernstein. Break it down: A comparison of macro-and microtasks. In *Proceedings of CHI*, 2015.
- [57] Lydia B. Chilton and Jaime Teevan. Addressing people’s information needs directly in a web search result page. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 27–36, New York, NY, USA, 2011. ACM.
- [58] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popovi, and Foldit players.

- Predicting protein structures with a multiplayer online game. *Nature Letter*, 466(7307):756760, 2010.
- [59] Justin Cranshaw and Aniket Kittur. The polymath project: Lessons from a successful online collaboration in mathematics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1865–1874, New York, NY, USA, 2011. ACM.
- [60] Tathagata Das, Prashanth Mohan, Venkata N. Padmanabhan, Ramachandran Ramjee, and Asankhaya Sharma. Prism: Platform for remote sensing using smartphones. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, MobiSys '10, pages 63–76, New York, NY, USA, 2010. ACM.
- [61] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255, June 2009.
- [62] Sonalde Desai, Amaresh Dubey, Brijlal Joshi, Mitali Sen, Abusaleh Shariff, and Reeve D. Vanneman. *Human Development in India: Challenges for a Society in Transition*. Oxford University Press, 2010.
- [63] Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts. An experimental study of search in global social networks. *Science*, 301:827–829, 2003.
- [64] Steven P. Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L.

Schwartz, and Scott R. Klemmer. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Trans. Comput.-Hum. Interact.*, 17(4):18:1–18:24, December 2010.

- [65] Nathan Eagle. Txteagle: Mobile crowdsourcing. In *Proceedings of the 3rd International Conference on Internationalization, Design and Global Development: Held As Part of HCI International 2009*, IDGD ’09, pages 447–456, Berlin, Heidelberg, 2009. Springer-Verlag.
- [66] Darren Edge, Stephen Fitchett, Michael Whitney, and James Landay. Memreflex: Adaptive flashcards for mobile microlearning. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI ’12, pages 431–440, New York, NY, USA, 2012. ACM.
- [67] Anthony Fader, Stephen Soderland, and Oren Etzioni. Identifying relations for open information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP ’11, pages 1535–1545, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.
- [68] Christopher M. Ford. Twitter, facebook and ten red balloons: Social network problem solving and homeland security. *HOMELAND SECURITY AFFAIRS*, 7(3), 2011.
- [69] Jon Froehlich, Tawanna Dillahunt, Predrag Klasnja, Jennifer Mankoff, Sunny Consolvo, Beverly Harrison, and James A. Landay. Ubigreen: Investigating a mo-

bile tool for tracking and supporting green transportation habits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1043–1052, New York, NY, USA, 2009. ACM.

- [70] Snehal (Neil) Gaikwad, Durim Morina, Rohit Nistala, Megha Agarwal, Alison Cossette, Radhika Bhanu, Saiph Savage, Vishwajeet Narwal, Karan Rajpal, Jeff Regino, Aditi Mithal, Adam Ginzberg, Aditi Nath, Karolina R. Ziulkoski, Trygve Cossette, Dilrukshi Gamage, Angela Richmond-Fuller, Ryo Suzuki, Jeerel Herrejón, Kevin Le, Claudia Flores-Saviaga, Haritha Thilakarathne, Kajal Gupta, William Dai, Ankita Sastry, Shirish Goyal, Thejan Rajapakshe, Niki Abolhassani, Angela Xie, Abigail Reyes, Surabhi Ingle, Verónica Jaramillo, Martin Godínez, Walter Ángel, Carlos Toxtli, Juan Flores, Asmita Gupta, Vineet Sethia, Diana Padilla, Kristy Milland, Kristiono Setyadi, Nuwan Wajirasena, Muthitha Batagoda, Rolando Cruz, James Damon, Divya Nekkanti, Tejas Sarma, Mohamed Saleh, Gabriela Gongora-Svartzman, Soroosh Bateni, Gema Toledo Barrera, Alex Peña, Ryan Compton, Deen Aariff, Luis Palacios, Manuela Paula Ritter, Nisha K.K., Alan Kay, Jana Uhrmeister, Srivalli Nistala, Milad Esfahani, Elsa Bakiu, Christopher Diemert, Luca Matsumoto, Manik Singh, Krupa Patel, Ranjay Krishna, Geza Kovacs, Rajan Vaish, and Michael Bernstein. Daemo: A self-governed crowdsourcing marketplace. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15 Adjunct, pages 101–102, New York, NY, USA, 2015. ACM.

- [71] Huiji Gao, Geoffrey Barbier, and Rebecca Goolsby. Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems*, 26(3):10–14, 2011.
- [72] Aakar Gupta, William Thies, Edward Cutrell, and Ravin Balakrishnan. mclerk: Enabling mobile crowdsourcing in developing regions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’12, pages 1843–1852, New York, NY, USA, 2012. ACM.
- [73] Severin Hacker and Luis von Ahn. Matchin: Eliciting user preferences with an online game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’09, pages 1207–1216, New York, NY, USA, 2009. ACM.
- [74] Kotaro Hara, Vicki Le, and Jon Froehlich. Combining crowdsourcing and google street view to identify street-level accessibility problems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’13, pages 631–640, New York, NY, USA, 2013. ACM.
- [75] Joseph M. Hellerstein and David L. Tennenhouse. Searching for jim gray: A technical overview. *Commun. ACM*, 54(7):77–87, July 2011.
- [76] Vaughn Hester, Aaron Shaw, and Lukas Biewald. Scalable crisis relief: Crowd-sourced sms translation and categorization with mission 4636. In *Proceedings of the First ACM Symposium on Computing for Development*, ACM DEV ’10, pages 15:1–15:7, New York, NY, USA, 2010. ACM.

- [77] Raphael Hoffmann, Saleema Amershi, Kayur Patel, Fei Wu, James Fogarty, and Daniel S. Weld. Amplifying community content creation with mixed initiative information extraction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1849–1858, New York, NY, USA, 2009. ACM.
- [78] Todd S. Braver B. J. Casey David Servan-Schreiber Jonathan D. Cohen, Steven D. Forman and Douglas C. Noll.
- [79] W James Kent, Charles W Sugnet, Terrence S Furey, Krishna M Roskin, Tom H Pringle, Alan M Zahler, and David Haussler. The human genome browser at ucsc. *Genome research*, 12(6):996–1006, 2002.
- [80] Shashank Khanna, Aishwarya Ratan, James Davis, and William Thies. Evaluating and improving the usability of mechanical turk for low-income workers in india. In *Proceedings of the First ACM Symposium on Computing for Development*, ACM DEV '10, pages 12:1–12:10, New York, NY, USA, 2010. ACM.
- [81] Joy Kim, Justin Cheng, and Michael S. Bernstein. Ensemble: Exploring complementary strengths of leaders and crowds in creative collaboration. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 745–755, New York, NY, USA, 2014. ACM.
- [82] Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. The future of crowd

- work. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, pages 1301–1318, New York, NY, USA, 2013. ACM.
- [83] Daphne Koller, Andrew Ng, Chuong Do, and Zhenghao Chen. Retention and intention in massive open online courses: In depth. *Educause Review*, 48(3):62–63, 2013.
- [84] Ponnurangam Kumaraguru and Lorrie Cranor. Privacy in india: Attitudes and awareness. In George Danezis and David Martin, editors, *Privacy Enhancing Technologies*, volume 3856 of *Lecture Notes in Computer Science*, pages 243–258. Springer Berlin Heidelberg, 2006.
- [85] Walter S. Lasecki, Christopher D. Miller, and Jeffrey P. Bigham. Warping time for more effective real-time crowdsourcing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2033–2036, New York, NY, USA, 2013. ACM.
- [86] Jeehyung Lee, Wipapat Kladwang, Minjae Lee, Daniel Cantu, Martin Azizyan, Hanjoo Kim, Alex Limpaecher, Snehal Gaikwad, Sungroh Yoon, Adrien Treuille, Rhiju Das, and EteRNA Participants. Rna design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences*, 111(6):2122–2127, 2014.
- [87] Chien-Wei Lin, Kuan-Ta Chen, Ling-Jyh Chen, Irwin King, and Jimmy HM Lee. An analytical approach to optimizing the utility of esp games. In *Web Intelligence*

and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on, volume 1, pages 184–187. IEEE, 2008.

- [88] Yarun Luon, Christina Aperjis, and BernardoA Huberman. Rankr: A mobile system for crowdsourcing opinions. In JoyYing Zhang, Jarek Wilkiewicz, and Ani Nahapetian, editors, *Mobile Computing, Applications, and Services*, volume 95 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 20–31. Springer Berlin Heidelberg, 2012.
- [89] Yarun Luon, Christina Aperjis, and BernardoA Huberman. Rankr: A mobile system for crowdsourcing opinions. *Springer Berlin Heidelberg*, 95:20–31, 2012.
- [90] Kurt Luther, Casey Fiesler, and Amy Bruckman. Redistributing leadership in online creative collaboration. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, pages 1007–1022, New York, NY, USA, 2013. ACM.
- [91] Thomas W. Malone and Michael S. Bernstein. *Handbook of Collective Intelligence*. MIT Press, 2015.
- [92] Winter Mason and Duncan J. Watts. Financial incentives and the "performance of crowds". In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '09, pages 77–85, New York, NY, USA, 2009. ACM.
- [93] Mary Meeker. *Internet Trends 2014*. Kleiner Perkins Caufield Byers, May 2014.

- [94] Eugene V Nalimov, Christoph Wirth, Guy McCrossan Haworth, et al. Kqqkqq and the kasparov-world game. *ICGA Journal*, 22(4):195–212, 1999.
- [95] Victor Naroditskiy, Iyad Rahwan, Manuel Cebrian, and Nicholas R. Jennings. Verification in referral-based crowdsourcing. *PLoS ONE*, 7(10), 2012.
- [96] Prayag Narula, David Rolnitzky, and Bjoern Hartmann. Mobileworks: A mobile crowdsourcing platform for workers at the bottom of the pyramid.
- [97] M. Nebeling, A. Guo, K. Murray, A. Tostengard, A. Giannopoulos, M. Mihajlov, S. Dow, J. Teevan, and J. P. Bigham. Wearwrite: Orchestrating the crowd to complete complex tasks from wearables (we wrote this paper on a watch). *arXiv*, 1508.02982, July 2015.
- [98] Ory Okolloh. Ushahidi, or ‘testimony’: Web 2.0 tools for crowdsourcing crisis information. *Participatory learning and action*, 59(1):65–70, 2009.
- [99] Eyal Ophir, Clifford Nass, and Anthony D. Wagner. Cognitive control in media multitaskers. *PNAS*, 106(37):15583–15587, 2009.
- [100] Antti Oulasvirta, Sakari Tamminen, Virpi Roto, and Jaana Kuorelahti. Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’05, pages 919–928, New York, NY, USA, 2005. ACM.
- [101] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: bringing order to the web. 1999.

- [102] Gabriele Paolacci, Jesse Chandler, and Panagiotis G. Ipeirotis. Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5(5), 2010.
- [103] Galen Pickard, Wei Pan, Iyad Rahwan, Manuel Cebrian, Riley Crane1, Anmol Madan, and Alex Pentland. Time-critical social mobilization. *Science*, 334(6055):509–512, 2011.
- [104] Ruth Pordes, Don Petrasick, Bill Kramer, Doug Olson, Miron Livny, Alain Roy, Paul Avery, Kent Blackburn, Torre Wenaus, Frank Würthwein, et al. The open science grid. In *Journal of Physics: Conference Series*, volume 78, page 012057. IOP Publishing, 2007.
- [105] Amy R. Poteete, Marco A. Janssen, and Elinor Ostrom. *Working Together: Collective Action, the Commons, and Multiple Methods in Practice*. Princeton University Press, 2009.
- [106] Moo-Ryong Ra, Bin Liu, Tom F. La Porta, and Ramesh Govindan. Demo: Medusa: A programming framework for crowd-sensing applications. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, MobiSys ’12, pages 481–482, New York, NY, USA, 2012. ACM.
- [107] 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.

- [108] Iyad Rahwan, Sohan Dsouza, Alex Rutherford, Victor Naroditskiy, James McInerney, Matteo Venanzi, Nicholas R. Jennings, and Manuel Cebrian. Global manhunt pushes the limits of social mobilization. *Computer*, 46(4):68–75, 2003.
- [109] Nimmi Rangaswamy and Edward Cutrell. Re-sourceful networks: Notes from a mobile social networking platform in india. *Pacific Affairs*, 85(3):587–606, 2012-09-01T00:00:00.
- [110] J Reich. MOOC completion and retention in the context of student intent. *ED-UCAUSE Review Online*, 2014.
- [111] Joel Ross, Lilly Irani, M. Six Silberman, Andrew Zaldivar, and Bill Tomlinson. Who are the crowdworkers?: Shifting demographics in mechanical turk. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 2863–2872, New York, NY, USA, 2010. ACM.
- [112] Lee Ross and Richard Nisbett. *The person and the situation: Perspectives of social psychology*. Pinter & Martin, 2011.
- [113] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.
- [114] Alex Rutherford, Manuel Cebrian, Sohan Dsouza, Esteban Moro, Alex Pentland,

- and Iyad Rahwan. Limits of social mobilization. *Proceedings of the National Academy of Sciences*, 110(16):6281–6286, 2013.
- [115] Niloufar Salehi, Lilly C. Irani, Michael S. Bernstein, Ali Alkhateeb, Eva Ogbe, Kristy Milland, and Clickhappier. We are dynamo: Overcoming stalling and friction in collective action for crowd workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI ’15, pages 1621–1630, New York, NY, USA, 2015. ACM.
- [116] C. Schuster, Bipeng Zhang, R. Vaish, P. Gomes, J. Thomas, and J. Davis. RTI compression for mobile devices. In *Information Technology and Multimedia (ICIMU), 2014 International Conference on*, pages 368–373, Nov 2014.
- [117] Alok Shankar Mysore, Vikas S. Yaligar, Imanol Arrieta Ibarra, Camelia Simoiu, Sharad Goel, Ramesh Arvind, Chiraag Sumanth, Arvind Srikantan, Bhargav HS, Mayank Pahadia, Tushar Dobha, Atif Ahmed, Mani Shankar, Himani Agarwal, Rajat Agarwal, Sai Anirudh-Kondaveeti, Shashank Arun-Gokhale, Aayush Atttri, Arpita Chandra, Yogitha Chilukur, Sharath Dharmaji, Deepak Garg, Namana Gupta, Paras Gupta, Glincy Mary Jacob, Siddharth Jain, Shashank Joshi, Tarun Khajuria, Sameeksha Khillan, Sandeep Konam, Praveen Kumar-Kolla, Sahil Loomba, Rachit Madan, Akshansh Maharaja, Vudit Mathur, Bharat Munshi, Mohammed Nawazish, Venkata Neehar-Kurukunda, Venkat Nirmal-Gavarraju, Sonali Parashar, Harsh Parikh, Avinash Paritala, Amit Patil, Rahul Phatak, Mandar Pradhan, Abhilasha Ravichander, Krishna Sangeeth, Sreecharan Sankara-

narayanan, Vibhor Sehgal, Ashrith Sheshan, Suprajha Shibiraj, Aditya Singh, Anjali Singh, Prashant Sinha, Pushkin Soni, Bipin Thomas, Kasyap Varma-Dattada, Sukanya Venkataraman, Pukit Verma, and Ishan Yelurwar. Investigating the "wisdom of crowds" at scale. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15 Adjunct, pages 75–76, New York, NY, USA, 2015. ACM.

- [118] R Silberzahn and EL Uhlmann. Crowdsourced research: Many hands make tight work. *Nature*, 526(7572):189, 2015.
- [119] Jonathan Silvertown. A new dawn for citizen science. *Trends in ecology & evolution*, 24(9):467–471, 2009.
- [120] Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing annotations for visual object detection, 2012.
- [121] Michael Szpir. Clickworkers on mars. *American Scientist*, 90(3):226, 2002.
- [122] John C. Tang, Manuel Cebrian, Nicklaus A. Giacobe, Hyun-Woo Kim, Taemie Kim, and Douglas.
- [123] Dario Taraborelli, Daniel Mietchen, Panagiota Alevizou, and Alastair Gill. Expert participation on wikipedia: barriers and opportunities. *Wikimania*, August 2011.
- [124] Bill Tomlinson, Joel Ross, Paul Andre, Eric Baumer, Donald Patterson, Joseph Cornelius, Martin Mahaux, Syavash Nobarany, Marco Lazzari, Birgit Penzenstadler, Andrew Torrance, David Callele, Gary Olson, Six Silberman, Marcus Stünder,

Fabio Romancini Palamedì, Albert Ali Salah, Eric Morrill, Xavier Franch, Florian Floyd Mueller, Joseph 'Jofish' Kaye, Rebecca W. Black, Marisa L. Cohn, Patrick C. Shih, Johanna Brewer, Nitesh Goyal, Pirjo Nakkki, Jeff Huang, Nilufar Baghaei, and Craig Saper. Massively distributed authorship of academic papers. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 11–20, New York, NY, USA, 2012. ACM.

- [125] Rajan Vaish, James Davis, and Michael Bernstein. Crowdsourcing the research process. *Collective Intelligence*, 3, May 2015.
- [126] Rajan Vaish, Keith Wyngarden, Jingshu Chen, Brandon Cheung, and Michael S. Bernstein. Twitch crowdsourcing: Crowd contributions in short bursts of time. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 3645–3654, New York, NY, USA, 2014. ACM.
- [127] Aditya Vashistha, Rajan Vaish, Edward Cutrell, and William Thies. The who-dunit challenge: Mobilizing the crowd in india. In Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler, editors, *Human-Computer Interaction INTERACT 2015*, volume 9297 of *Lecture Notes in Computer Science*, pages 505–521. Springer International Publishing, 2015.
- [128] A. Veit, M. Wilber, R. Vaish, S. Belongie, J. Davis, V. Anand, A. Aviral, P. Chakrabarty, Y. Chandak, S. Chaturvedi, C. Devaraj, A. Dhall, U. Dwivedi, S. Gupte, S. N. Sridhar, K. Paga, A. Pahuja, A. Raisinghani, A. Sharma, S. Sharma, D. Sinha, N. Thakkar, K. Bala Vignesh, U. Verma, K. Abhishek,

- A. Agrawal, A. Aishwarya, A. Bhattacharjee, S. Dhanasekar, V. Karthik Gullapalli, S. Gupta, C. G, K. Jain, S. Kapur, M. Kasula, S. Kumar, P. Kundaliya, U. Mathur, A. Mishra, A. Mudgal, A. Nadimpalli, M. Sree Nihit, A. Periwala, A. Sagar, A. Shah, V. Sharma, Y. Sharma, F. Siddiqui, V. Singh, S. Abhinav, and A. D. Yadav. On Optimizing Human-Machine Task Assignments. *ArXiv e-prints/AAAI HCOMP Work-In-Progress*, September 2015.
- [129] Luis Von Ahn. Games with a purpose. *Computer*, 39(6):92–94, 2006.
- [130] Luis von Ahn. Duolingo: Learn a language for free while helping to translate the web. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, IUI ’13, pages 1–2, New York, NY, USA, 2013. ACM.
- [131] Luis Von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. reCAPTCHA: Human-based character recognition via web security measures. *Science*, 321(5895):1465–1468, 2008.
- [132] Steven Weber. *The success of open source*, volume 897. Cambridge Univ Press, 2004.
- [133] Susan P. Wyche, Andrea Forte, and Sarita Yardi Schoenebeck. Hustling online: Understanding consolidated facebook use in an informal settlement in nairobi. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’13*, pages 2823–2832, New York, NY, USA, 2013. ACM.
- [134] Tingxin Yan, Matt Marzilli, Ryan Holmes, Deepak Ganesan, and Mark Corner.

mCrowd: A Platform for Mobile Crowdsourcing. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, SenSys '09, pages 347–348, New York, NY, USA, 2009. ACM.

- [135] Chen-Hsiang Yu. Mobile continuous reading. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 1405–1410, New York, NY, USA, 2012. ACM.