5

Survey of Industry Users: Summary and Methodology

In this chapter, we begin describing our survey of industry users of crowd-sourcing. We begin with an executive high-level summary of our findings (Section 5.1), discuss our survey methodology (Section 5.2), and describe three categories of industry users we identified (Section 5.3).

Over the next few chapters, we describe the detailed survey results, specifically:

- STATISTICS: the statistics of the crowdsourcing deployments (Section 6),

- USE CASES: the applications crowdsourcing is typically used for, approaches adopted before the advent of crowdsourcing, and the benefits provided by crowdsourcing (Chapter 7),

- DETAILS: ensuring quality, incentivization, task design, and other deployment challenges. (Chapter 8)

We begin with a high level summary of our findings; given the extent of our undertaking, it is impossible to summarize all of our takeaways into one section. We therefore encourage even the impatient reader to at least skim all four chapters, with a special eye towards the excerpted quotes and tables; the

former containing special anecdotes, use-cases, or lessons learned, the latter containing summaries of the findings of the sections within the chapters.

5.1 Executive summary

Crowdsourcing is common. Crowdsourcing is alive and well at the companies we spoke with: none are indicating that they are decreasing their investment in it, and almost all of them were looking for new ways to use it.

Crowdsourcing deployments are large-scale. The statistics on crowd management and the teams that build crowd-powered systems were illuminating (Section 6). At their largest, participants reported hundreds of employees deploying hundreds of thousands of tasks per week, with overall spending in the millions of dollars per year. The number of employees building crowdsourcing-oriented tools ranged from 1 company-wide to "tens to hundreds" at larger organizations. The participant with the highest paid task throughput reported processing about 400,000 tasks per week, and the most popular response across participants was in the low tens of thousands of tasks per week. At the low end, participants reported spending $300-$1000 per week. The two largest participants that provided us with numbers reported spending approximately $10,000 and $30,000 per week respectively.

Many users host their own platforms, with long worker relationships. Five participants hosted their own crowd work platforms (i.e., they use an intermediary or outsourcing company to hire workers who work on tasks provided by the participant 9–5). The ubiquity of internal crowd work platforms was one of the most surprising findings from our study; indicating that academic research, which is focused on popular platforms like CrowdFlower and Mechanical Turk, has summarily ignored one of the most common industry mechanisms for employing crowdsourcing. Apart from the five that only use their internal platform, five utilized an external provider's platform, and two did a mix of both. When participants hosted their own platform, the length of relationship with workers was high: The most common response for maximum worker tenure was 3 years, with medians between 1 and 2.5 years. At the low end, participants interacted with 50-100 workers per week, and at the high end, participants interacted with 100s to low 1000s of workers in a week.

Many new and novel uses of crowds. Some companies are doing interest-

ing and novel things with the crowd (Section 7.1). For example, one participant has such a trusting relationship with a few hundred crowd workers that the participant feels comfortable pre-paying the workers to monitor the news for updates on companies of interest and updates crowd workers' balances as they send back "new" facts. Yet another participant uses crowdsourcing frameworks on in-house employees, who work on tasks whenever they get time free from their regular work.

Classification and entity resolution are most popular uses of crowds. While there are many interesting use cases, most of the ones participants described are relatively standard. The two most popular use cases are classification and entity resolution. There are often large teams that use crowdsourcing for only one targeted application (e.g., categorization, or data extraction); they have spent many months tuning their deployment for this application, and use it periodically.

Most problems solved by crowdsourcing are unsolvable without it. We wanted to understand how crowdsourcing was perceived at various companies. When asked how they solved problems before the advent of crowd work, a little less than half of the responses were of the form "our company didn't exist at that point," or "we didn't solve this problem before crowdsourcing" (Section 7.2). When we asked participants to explain some of the benefits of crowd work, the top three responses pointed to the flexibility to scale work up and down, the low cost, and that crowdsourcing enabled previously difficult or impossible tasks (Section 7.3). One participant told us that the main benefit he derives from crowd work is that he doesn't need to "argue with management" to hire the manpower to get things done; he can simply use small amounts of money as the need arises.

Quality management schemes are somewhat primitive. Most participants use very simple schemes, such as majority vote over multiple worker responses, to remove errors. However, more than half of the participants do use some form of simple Expectation-Maximization scheme to reason about worker error rates and task answer quality (Section 8.1). Most participants do not do any optimization to reduce cost while keeping accuracy fixed.

Incentivization schemes are primitive. The most common methods for incentivizing workers were financial: per-task payment was the most popular,

followed by bonuses and hourly payments (Section 8.2). Less tangible incentives, like gamification, leaderboards, and promotions, were less popular.

Industry users rarely use workflows/toolkits from academia. Finally, while we have so far described how participants successfully used crowd work, researchers might also wonder how some of the more complex approaches from the literature have fared in practice (Section 8.3). Surprisingly, no participants utilized third-party frameworks to simplify their crowd-powered data processing workflows. Many participants reported that none of their workflows have more than crowdsourcing step, suggesting that participants are looking for simple tasks to be completed rather than the more in-depth multi-stage workflows perscribed by crowdsourcing researchers. In support of this observation, less than a third of participants claimed to use at least one crowdsourcing "design" pattern (iterative refinement, find-fix-verify, or do-verify—we will describe these later) from the literature.

As you read this chapter, watch out for various tables and pull quotes to guide you through our findings. Much of the summary above has been extracted from these pull quotes, and we hope they can serve as guideposts for your reading.

5.2 Survey and recruitment methodology

Here, we describe our survey methodology for both the survey of industry users, as well as the survey in Chapter 9 of marketplace providers.

With a goal of identifying the key use cases, existing solutions, and open research problems in the field of crowd-powered data processing, we conducted surveys of two groups of stakeholders: 1) industry users of crowdsourcing, and 2) operators of crowd labor marketplaces. Each survey consisted of approximately 45 minutes worth of questions. Participants had an option of taking the survey synchronously by phone or asynchronously over email. When a participant requested a phone interview, the authors took notes and coded the responses on behalf of the participant.

To generate the surveys, we first created a list of topic areas we wished to learn more about, and then iteratively generated questions, recategorizing questions into topic areas as appropriate. We received feedback on the questions from an expert in survey generation. For the industry survey, we piloted

Description of Survey Section	# Questions	Paraphrased Example Questions	Chapter
Crowd Use Cases	4	— Which of the following use cases of crowds apply to the tasks your team is solving? (examples include classification, text generation, etc.) — How did you solve these problems before crowdsourcing?	7
Crowd Management Statistics	6	— How many people in your organization work on crowdsourcing? — How many tasks per week do you generate?	6
Quality of Work and Workers	5	— How do you evaluate worker quality? — Do you provide feedback to workers?	8
Incentives/Payment Mechanisms	3	— Do you pay workers hourly or per task? —Are there different classes/tiers of workers?	8
Task Design/Decomposition	7	— What crowd management frameworks do you use? — Do you primarily create microtasks or macrotasks?	8

Table 5.1: A summary of the types of questions we asked participants either through a survey they filled out on their own time or through phone interviews. The example questions provided are paraphrased descriptions. Detailed questions can be found in Appendix A.

the survey with one participant by voice and one by email and further clarified the questions based on confusion that arose during the interviews, but kept the responses from the two pilot participants. There were too few marketplace participants (4) for a pilot phase, but we conducted the marketplace surveys after the industry ones, and were able to clarify marketplace questions from that experience.

The industry survey consisted of six sections: use cases, infrastructure for managing workers, tools for inferring worker and work quality, incentive and payment mechanisms, task design and decomposition, and a ranking of challenges. The marketplace survey covered crowd demographics, descriptions and summary statistics of common implementations/use cases, and worker and work quality assurance. The chapters that cover the results of each section of the survey are also listed. Table 5.1 describes each section in more detail, and the industry and marketplace surveys can respectively be found in Appendices A and B.

To recruit participants, we surveyed our own social networks and the crowdsourcing literature (conferences, workshops, and blog posts) for industry participants, and contacted them for their participation. To expand the

Company	Team	Persona
Amazon	Product classification	Largely single-case user
Captricity	Focus of large part of company	Largely single-case user
Dropbox	Single person consulting several teams	Multi-case user / Internal provider
Facebook	Entities team	Multi-case user
Flipora	Startup CTO	Multi-case user
GoDaddy	Small business data extraction	Multi-case user
Groupon	Merchant data team	Multi-case user
Google	Internal crowdsourcing team	Internal provider
Google	Web knowledge discovery team	Multi-case user
LinkedIn	Single person consulting several teams	Multi-case user / Internal provider
Microsoft	Internal crowdsouricng team	Internal provider
Microsoft	Search relevance team	Multi-case user
Youtube	Crowdsourcing team	Largely single-case user

Table 5.2: A summary of the company and persona of the team that we spoke with in that company. Some organizations (e.g., Microsoft, Google) are so large that we were able to speak with both a multi-case user and an internal provider. Note that some teams (e.g., Dropbox, LinkedIn) were largely composed of a single person that both implemented crowdsourcing solutions and consulted other teams on crowdsourced implementations.

scope of the survey, we asked that initial set of participants for any relevant connections that they had. Participants were ensured that their responses would only be reported in aggregate, except for meaningful quotes that they would be allowed to review. Table 5.2 identifies our survey participants.

5.3 Three personas of industry users

After evaluating responses to the industry user survey, we identified three team personas that we later use to summarize some of our findings. While these personas don't always utilize crowd work similarly, their behaviors of teams with the same persona are often similar.

Internal providers (4/13). These teams serve as tool- and service-builders for other crowdsourcing users within their company. They are often the go-to team that provides consulting in addition to the tools that they build. As a sign of the breadth of their experience, three of the four internal providers we surveyed saw every data processing use case that we listed. These specialized intermediary teams are more common in larger companies with varied needs.

Multi-case users (8/13). These teams directly solve problems with crowd workforces, rather than serving as intermediaries as the internal providers do. Forming the largest set of participants by far, these teams do not see as many use cases as the internal providers, but have end-to-end experience solving several problems for their companies.

Largely single-case users (3/13). These users use crowds in a small number of workflows for primarily one task within their company or team, and can often reflect on several iterations of their solutions to this one problem. The small number of use cases should not be conflated with less experience with crowd work: one of our largely single-case participants consistently generates hundreds of thousands of tasks per week, amongst the largest task generation volume of any participant.

Some organizations (e.g., Microsoft, Google) are so large that we were able to speak with both a multi-case user and an internal provider. Note that some teams (e.g., Dropbox, LinkedIn) were largely composed of a single person that both implemented crowdsourcing solutions and consulted other teams on crowdsourced implementations.