

The scientific state of the art in crowdsourcing

An overview of crowdsourcing, its different facets, challenges and criticisms

Diego Barbera
1327823

Viktor Kvaternjak
1328111

Andreas Ntaflos
0326302

Richard Plangger
1025637

Christian Schnitzer
0828380

Konrad Steiner
0927159

ABSTRACT

Crowdsourcing as a practice for outsourcing work to an unidentified public is relatively young. In this paper we review current scientific literature for an overview of crowdsourcing, focusing on a consistent definition of the concept and differentiating between the two main forms of crowdsourcing initiatives: those that are creative or intellectual in nature, rewarding unique ideas and solutions to problems, and those that involve solving so-called micro-tasks that are impossible for a computer but easy for human intelligence. These two forms of crowdsourcing involve very different types of crowds with different motivations and behaviour. Whereas creative or intellectual initiatives involve often rich financial rewards, building of communities, practicing special skills and passion for the activity, the crowds of micro-task initiatives are anonymous, culturally and geographically diverse, and tasks pay only pennies. These initiatives need to be designed in a way that prevents cheating and manipulation while ensuring quality of work output.

1. INTRODUCTION

The term *crowdsourcing* has been coined by Jeff Howe in 2006 [9] as a portmanteau of *crowd* and *outsourcing*. He describes it generally as outsourcing work to an “undefined, generally large group of people in the form of an open call” [10]. The term and the underlying practice are still quite young and the theoretical backgrounds—What is a crowd? What motivates the crowd? Why do some crowdsourcing systems and initiatives thrive while others fail miserably?—are still being researched, for example in the works of Daren Brabham [2, 3, 4]. This is why in scientific literature we find various, often directly conflicting definitions of crowdsourcing, as well as conflicting classifications of existing systems and initiatives; some authors see Wikipedia or Youtube as crowdsourcing examples while others argue the exact opposite [7].

In this paper we give an overview of crowdsourcing by discussing an integrated definition of the term and practice

as put forth by Estellés-Arolas and González-Ladrón-de-Guevara in [7] and examining current crowdsourcing systems and platforms based on that definition. We also compare the motivations and composition of different kinds of crowds for different kinds of problems and how these crowds often differ widely in these regards. Next we describe broadly how to design a crowdsourcing initiative to be successful and which steps can be taken to control reliability and quality of work. Finally we discuss some critical aspects of and problems with crowdsourcing which are often overlooked or ignored in current literature.

2. CROWDSOURCING DEFINED

The most comprehensive definition of crowdsourcing has been compiled by Estellés-Arolas and González-Ladrón-de-Guevara. In their work [7] the authors analysed a large number of existing definitions and extracted various characteristics that apply to a crowdsourcing system, resulting in the following:

Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage that what the user has brought to the venture, whose form will depend on the type of activity undertaken.

Based on that definition all of the following characteristics, as identified in [7] apply to a crowdsourcing system:

- (a) There is a clearly defined crowd
- (b) There exists a task with a clear goal
- (c) The recompense received by the crowd is clear
- (d) The crowdsourcer is clearly identified
- (e) The compensation received by the crowdsourcer is clearly defined
- (f) It is an online assigned process of participative type
- (g) It uses an open call of variable extent

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

- (h) It uses the internet

Having identified the characteristics that describe a crowdsourcing system it is now easy to see that neither Wikipedia, nor Youtube qualify as such. For Wikipedia the missing characteristics are (d), (e) and (g), while for Youtube none except (a) and (h) even apply.

The reason crowdsourcing can and often does work is based on the concept of a “wisdom of crowds”, as defined and discussed by James Surowiecki in [17]. Based on various case studies Surowiecki argues that under the right circumstances “groups can be smarter than the smartest person in them” and that the “wisdom of crowds” is derived from aggregating solutions, not by averaging them. Touching upon these insights Brabham confirms that the internet provides “the perfect technology” to aggregate independent and disparate solutions [2] conceived by the crowd.

3. OVERVIEW OF CROWDSOURCING SYSTEMS AND PLATFORMS

Modern crowdsourcing systems in general can be divided into three different types, as discussed in [11]: aggregator platforms, specialised platforms and crowd provider platforms.

Aggregator platforms such as Crowdfunder¹ and Crowdsource² primarily focus on developing crowd-based solutions for businesses that have existing workflows in need of crowdsourcing. Such platforms do not maintain their own crowd or workforce but recruit workers from specialised or crowd provider platforms and thus provide a high level of abstraction from crowdsourcing issues like quality control and recruiting.

Specialised platforms such as Threadless³, InnoCentive⁴ or Microtask⁵ manage and maintain their own workforce and focus on very specific tasks, such as t-shirt design, solving R&D problems or document processing and data entry. As such the workforce is usually limited in variety, location, expertise and other aspects.

Crowd provider platforms such as Amazon Mechanical Turk⁶, Clickworker⁷ or Microworkers⁸ are the most flexible of the three crowdsourcing system types, as they have the largest and most diverse crowd of workers that can be directly accessed, and allow platform employers to design and implement a large variety of tasks via web interface or API with almost no restrictions. Typical tasks solved on crowd provider platforms are image tagging, reviewing and correcting OCR results or transcribing audio to text. Such tasks are easily solved using human intelligence but usually impossible for a computer.

Of these three types of platforms we focus on the *specialised* and on the *crowd provider* platforms. The aggregator platform can be seen as a sort of “meta” platform, simply making good use of the latter two and shielding the platform employer

from the more intricate details of designing and running a crowdsourcing initiative.

As discussed in section 2 there must exist a clearly defined task for a clearly defined crowd to solve. Tasks on specialised platforms can differ greatly from tasks on crowd provider platforms. The former almost invariably require creativity, expertise and time to solve, while the latter are mostly repetitive and easily solved in mere seconds or minutes without much cognitive effort. These are called *micro-tasks*.

The crowd’s recompense for solving creative or R&D tasks is usually vastly greater than that for solving micro-tasks. In the case of InnoCentive the prize for accepted solutions can be as high as \$100 000 or in special cases even \$1 M. In contrast the reward for solving a single micro task is usually notably less than \$1.

From the above it can be concluded that not every crowdsourcing initiative, workflow or campaign can and will work on every platform, and that different initiatives require different kinds of tasks, solved by different kinds of crowds.

4. MOTIVATING DIFFERENT CROWDS FOR DIFFERENT PROBLEMS

As discussed in the previous section it is obvious that different kinds of problems attract different kinds of crowds. Thus the motivation and make-up of crowds that solve creative or intellectual tasks is very different from crowds that work on non-demanding micro-tasks. Questions about motivation and composition of crowds are subject of ongoing research [1, 3, 4, 6, 8, 12, 14].

In the following paragraphs we discuss Brabham’s analysis of motivation and make-up of the crowds at iStockphoto⁹ [3] and Threadless [4], as two examples of crowdsourcing initiatives involving highly creative tasks.

As explained in [3] at iStockphoto the crowd, whose members are called “iStockers”, creates and uploads stock photographs, animations or videos. Clients find and download the stock they want, and individual photographers make a small profit for each download, with iStockphoto taking a portion of each profit. In accordance with the integrated definition [7] discussed in section 2 iStockphoto is indeed a crowdsourcing initiative: the clearly defined problem is producing stock photography which is put into the form of an open call on the internet, with solutions provided by a crowd of (amateur) photographers. Both the crowd and the crowdsourcer profit from clients purchasing the created stocks.

Brabham finds that the crowd at iStockphoto is elite and relatively homogenous, and even though many nationalities are represented the dominant group consists of middle- to upper-class, higher educated whites that are heavy internet users with high-speed internet connections at home. Regarding their motivation he finds that making money is the most important factor for members of the iStockphoto crowd. Peer recognition, learning new skills and the joy of participation are only secondary motivators. He notes that interestingly crowd participants at iStockphoto are generally not interested in the online community aspect and do not find their motivation in building a network of friends and creative professionals through the site.

Analysing the crowd at Threadless in [4] Brabham comes to similar, yet different conclusions. Threadless is “an online

¹<http://crowdfunder.com>

²<http://www.crowdsource.com>

³<http://www.threadless.com>

⁴<http://www.innocentive.com>

⁵<http://www.microtask.com>

⁶<http://www.mturk.com>

⁷<http://www.clickworker.com>

⁸<http://www.microworkers.com>

⁹<http://www.istockphoto.com>

t-shirt company that crowdsources the design process for its shirts through an ongoing online competition where registered members on the site can rate the design submissions of fellow members on a five-point scale.” [4] The highest-rated shirts are then produced and sold on the site, and the winning designers are rewarded with \$2000 in cash and \$500 in Threadless gift certificates. From this we can see that the integrated definition of crowdsourcing [7] applies to Threadless as well.

The biggest motivator at Threadless, Brabham finds, is the opportunity to make money, much like at iStockphoto, with developing and honing creative skills and leveraging freelance design work being secondary motivators. But at Threadless the love of community is an important motivator that is missing at iStockphoto. Another aspect unique to Threadless is that of crowd participants being “addicted” to the site and the community, meaning they see themselves as more as meaningful actors in the company’s business process than mere customers. It is this kind of “obsessive” and vibrant community that makes Threadless such an exemplary and successful crowdsourcing initiative.

In contrast to these creative crowdsourcing types stand micro-task-oriented initiatives found at Amazon Mechanical Turk (AMT) and other crowd provider platforms. In the following paragraphs we discuss research on crowd motivation based on studies using AMT.

At AMT the crowd consists of over 100 000 workers from over 100 countries [5] and is thus much larger and more diverse than crowds at specialised platforms like discussed above. These workers solve micro-tasks (so-called *Human Intelligence Tasks*, or HITs) as posted by requesters, for which they are paid upon completion. The reward for each task is set by the requester upon creation and is usually no more than a few cents. Requesters may refuse to pay workers who submit low-quality work, which may make workers with high refusal rates ineligible for participation in other tasks. Amazon as crowd provider charges 10%. Again the integrated crowdsourcing definition [7] applies.

Research by Horton [8], Kaufmann [12] and Mason [14] focuses mainly on how work output quality and quantity vary when varying the financial reward for solving tasks, while Chandler [6] examines how the perceived meaningfulness of a task affects quality and quantity of output.

According to Mason [14] increasing payments for tasks leads to increased quantity of work output, but not quality. Horton [8] finds complementary that workers work less when compensation is lower but do not work less when the task is more time consuming.

The authors of [14] also find that the design of the compensation scheme (e.g., quota versus piece rate) can significantly affect the quality of work output, to a point where better work can be produced for less pay. This may even result in high-quality work for no financial compensation, when the task’s intrinsic motivation is big enough.

This conclusion is supported by Chandler [6]. Here the authors experimented with the perceived meaningfulness of tasks: assisting in cancer research makes a task more meaningful, while knowing that the work output will be discarded makes a task less meaningful. The authors find that higher meaning increases quantity of work output (as well as quality, but only insignificantly) and requires less compensation, and that, interestingly, tasks with lower meaning lead to a decrease in quality of work output, but not in quantity.

From this discussion on the motivation and make-up of different crowds it can be seen that the quality of work output is more easily recognized or confirmed at crowdsourcing platforms that offer creative or intellectual tasks. They are, by their nature, a meritocracy and it is easy to “pick out a winner” based on the best-voted t-shirt design or the best-selling stock image. On the other hand it is much more difficult to validate the quality of solutions of micro tasks. Simply manually checking the entirety of every worker’s output would defeat the purpose of crowdsourcing work and also be infeasible, especially in very large crowdsourcing initiatives. We discuss such challenges and possible solutions in the next section.

5. RELIABILITY IN CROWDSOURCING INITIATIVES

In micro-task crowdsourcing initiatives it is usually not possible to quickly and easily check the crowds work output for quality or correctness. This can lead workers to “game” the system and solve micro-tasks in a random or malicious fashion, with the intent to quickly produce as much work output as possible (and getting paid for it) without investing any real effort. Such behaviour has been observed in various experiments involving initiatives posted to Amazon Mechanical Turk [11, 13, 15]. The authors of these experiments have devised recommendations on designing a crowdsourcing initiative and its tasks in ways that limits cheating or makes it easier to spot nonsensical work output.

In [13] the authors crowdsourced the rating of Wikipedia articles and found three ways to help eliminating most cheating or malicious workers. First, the tasks should have some questions that are objectively verifiable, such as how many references a given article cites. Second, the tasks should be designed in a way that makes solving it honestly require as much as or less effort than cheating. Third, there should be multiple ways to spot suspect solutions that work even for subjective tasks, such as suspiciously short solving times. The first way especially signals workers that their work output can and will be scrutinised and the third way allows filtering obviously bogus work output early on.

Similar campaign design recommendations are found by the authors of [11, 15], who engaged in user studies concerning quality of experience which are highly subjective in nature, such as rating images on recognisability and aesthetic appeal. They find that a reliable crowd can be created by letting workers complete simple and verifiable test tasks first and only allowing those that passed to participate in the main task. Additionally the main task should contain verification tests, such as captchas, consistency tests, such as “In which country do you live?” and later “On which continent do you live”, and content questions about the task itself. The authors warn against adding too many such tests and questions, lest workers quit the task in frustration.

6. CRITICAL CONSIDERATIONS OF CROWDSOURCING

Crowdsourcing is viewed in quite a positive light in current literature which often ignores potential or real drawbacks, and doesn’t contain much in the way of critical discussion of the practice. Thus, before concluding we present a more critical view of crowdsourcing based on [16] and [2].

An often-cited pro-crowdsourcing argument is the diversity of the crowd and that it contains participants from around the world [9, 10]. However, as Brabham finds in [3] the crowd at iStockphoto consists mostly of upper or middle class, English speaking whites with higher education and high-speed internet connections. This is probably especially representative of crowds at other creative and intellectual crowdsourcing platforms, and Brabham states “If solutions are measured against the yardstick of the company sponsoring the crowdsourcing application, or measured against the opinions of the homogenous crowd, alternatives to the presiding discourse will probably always lose out.” Meaning the solutions provided by such a homogenous crowd are less “disruptive” and alternative because they originate from very similar socio-economical backgrounds.

Additional problems with crowdsourcing include the discussed possibilities of cheating, manipulation and gaming, as well as pranking and companies which crowdsource tasks denying responsibility for failures by simply blaming everything on “the crowd.” Another issue is that of labour exploitation. Companies crowdsourcing designs or R&D problems benefit hugely from clever solutions by the crowd while the winning crowd participants are rewarded with literally pennies on the dollar.

It might also be argued that crowdsourcing may drive professionals (such as stock photographers or designers) out of business, but that discussion about capitalism is best left for economists and sociologists. In the end the real winners of crowdsourcing are the companies running the crowd provider platforms, such as Amazon or Clickworker, who profit from every problem posted and every solution devised.

7. CONCLUSION

In this paper we provided a short overview of the practice of crowdsourcing based on current literature and its often inconsistent definitions. We reviewed the differences between creative/intellectual crowdsourcing initiatives and those based on micro-tasks requiring very little cognitive effort, and discussed research on the motivation and make-up of different crowds solving different problems. We also examined the recommendations authors of crowdsourcing experiments provide for designing a crowdsourcing campaign that is not easily cheated or gamed and closed with a short discourse on some critical consideration of the crowdsourcing practice which are often ignored in the literature.

8. REFERENCES

- [1] M. Antikainen, M. Mäkipää, and M. Ahonen. Motivating and supporting collaboration in open innovation. *European Journal of Innovation Management*, 13(1):100–119, 2010.
- [2] D. C. Brabham. Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence: the international journal of research into new media technologies*, 14(1):75–90, 2008.
- [3] D. C. Brabham. Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application. *First Monday*, 13(6), 2008.
- [4] D. C. Brabham. Moving the crowd at threadless: Motivations for participation in a crowdsourcing application. *Information, Communication & Society*, 13(8):1122–1145, 2010.
- [5] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon’s mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5, 2011.
- [6] D. Chandler and A. Kapelner. Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization*, 2013.
- [7] E. Estellés-Arolas and F. González-Ladrón-de Guevara. Towards an integrated crowdsourcing definition. *Journal of Information science*, 38(2):189–200, 2012.
- [8] J. J. Horton and L. B. Chilton. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM conference on Electronic commerce*, pages 209–218. ACM, 2010.
- [9] J. Howe. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4, 2006.
- [10] J. Howe. *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business*. Three Rivers Press, 2009.
- [11] T. Hoßfeld, A. Mäder, K. Tutschku, F.-U. Andersen, C. Kappler, H. de Meer, I. Dedinski, and J. Oberender. CrowdTesting: A Novel Methodology for Subjective User Studies and QoE Evaluation. Technical Report 486, University of Würzburg, Institute of Computer Science, February 2013.
- [12] N. Kaufmann, T. Schulze, and D. Veit. More than fun and money. worker motivation in crowdsourcing-a study on mechanical turk. In *AMCIS*, 2011.
- [13] A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 453–456. ACM, 2008.
- [14] W. Mason and D. J. Watts. Financial incentives and the performance of crowds. *ACM SigKDD Explorations Newsletter*, 11(2):100–108, 2010.
- [15] J. A. Redi, T. Hoßfeld, P. Korshunov, F. Mazza, I. Povoia, and C. Keimel. Crowdsourcing-based multimedia subjective evaluations: a case study on image recognizability and aesthetic appeal. In *Proceedings of the 2nd ACM international workshop on Crowdsourcing for multimedia*, pages 29–34. ACM, 2013.
- [16] H. Simula. The rise and fall of crowdsourcing? In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, pages 2783–2791. IEEE, 2013.
- [17] J. Surowiecki. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Doubleday, 2004.