Chapter 3
Crowdsourcing Technology to Support Academic Research

Matthias Hirth[1], Jason Jacques[2], Peter Rodgers[3], Ognjen Scekic[4], and Michael Wybrow[5]

3.1 Introduction

Many academic research studies have small numbers of participants. One reason for this is the difficulty of finding participants to take part in research, especially when people with certain characteristics are required. Most such research studies would welcome additional participants. As such, there is growing interest from researchers in the use of crowdsourcing platforms due to the large populations of workers.

Despite the diversity of current commercial crowdsourcing platforms, most of them lack of support for academic research and its special needs. In this chapter we discuss the possibilities for practical improvement of academic crowdsourced studies through adaption of technological solutions.

As of April 2016 Crowdsourcing.org[1] lists over 130 web sites focusing on crowd labour. Noticably absent from this list are large platforms like Witmart (formerly Zhubajie)[2] which itself has about 13 million users. Most of these commercial platform providers focus mainly on large scale requesters with repetitive types of microtasks. The special needs and the comparatively low number of tasks submitted by researchers make them unattractive as main business customers for most providers.

Chair of Communication Networks, University of Würzburg, Germany
e-mail: matthias.hirth@informatik.uni-wuerzburg.de · Department of Engineering, University of Cambridge, UK
e-mail: jtj21@cam.ac.uk · School of Computing, University of Kent, UK
e-mail: P.J.Rodgers@kent.ac.uk · Distributed Systems Group, TU Wien, Vienna, Austria
e-mail: o.scekic@dsg.tuwien.ac.at · Faculty of Information Technology, Monash University, Australia
e-mail: Michael.Wybrow@monash.edu

[1] Crowdsourcing LLC. "List of Crowdsourcing Providers". http://www.crowdsourcing.org/directory (accessed April 2016).

[2] ZBJ Network Inc. "Witmart". http://www.witmart.com (accessed April 2016).

Current crowdsourcing systems do not fully support scientific research as the requirements are often very different from common commercial use cases. While platforms like Prolific Academic[3] aim to fill this niche, they still fall short of providing many of the necessary features. Researchers will often try to overcome the limitations of a platform by designing specialised software tools, e.g., for crowdsourced Quality of Experience tests (see Chapter 6). However, as these software tools are only loosely coupled to the actual crowdsourcing provider they cannot compete with the full potential of a commercial crowdsourcing platform with integrated support for academic research.

This chapter is focussed on the needs of academic studies performed on commercial crowdsourcing systems. We do not include internal enterprise crowd systems, or other closed work allocation systems such as EasyChair, as these typically lack the flexibility necessary to be used for academic studies. We discuss existing platforms and propose enhanced features, many of which would be relatively easy to implement, that would greatly assist the adoption of crowdsourcing as a mechanism for academic study.

This chapter is organized into two main sections: Section 3.2 provides an overview of the current state-of-the-art in crowdsourcing technology, whilst Section 3.3 has a detailed discussion of possible new technology and features. This second section includes Subsection 3.3.1 on possible improvements to the user management of crowdsourcing platforms; Subsection 3.3.2 on technological solutions to payment issues; Subsection 3.3.3 on the ethical aspect of technology for crowdsourcing; Subsection 3.3.4 on further hardware and instrumentation that might be adopted for crowdsourced studies; Subsection 3.3.5 on the potential for advanced study designs provided for by technological improvements. Finally Section 3.4 gives our conclusions.

3.2 Existing crowdsourcing platforms

Crowdsourcing aims to leverage a huge and diverse set of people to efficiently solve tasks that cannot easily be solved computationally. This is made possible by online platforms providing tools for “requester” users to create microtasks and make these available to “worker” users. In the following section we give a brief overview of the basic functionality currently available in commercial crowdsourcing platforms, where workers are financially rewarded for completed microtasks. Non-commercial crowdsourcing approaches, like posting microtask on social networks or online communities, or platforms focusing on voluntary paticipation, e.g., *Galaxy Zoo* or *Zooniverse* [40,41], are not considered as the implementation effort required to develop such platforms means they are only applicable to large scale projects. Thereafter given this brief overview of the basic functionality, we present a coarse-grained categorisation

[3] Prolific Academic. “Prolific”. http://prolific.ac/ (accessed April 2016).

scheme for those platforms that helps to identify a suitable platform type for specific use cases. Finally, we discuss the suitability of current commercial systems for use in academic research.

3.2.0.1 Functions of a crowdsourcing platform

Crowdsourcing platforms act as mediator between workers and requesters. However, most platform operators focus more on providing features that benefit requesters, since they are the customers of the service. In general, crowdsourcing platforms aim to support requesters in three main aspects: (1) Managing the crowdsourcing workforce, (2) creation of the microtasks, and (3) processing of the microtasks.

Maintaining a large and diverse workforce is one of the key aspects in crowdsourcing but also one of the most challenging ones. One reason is that an equilibrium between requesters and workers is required. That is, enough microtasks need to be submitted to keep the workers active, and enough workers need to be available to complete the available microtasks within the time constraints of the requesters. Another reason is the complexity of the remuneration for international workers, due to the different banking systems and legal constraints. Both aspects are completely abstracted for a requester on a commercial crowdsourcing platform. Moreover, some crowdsourcing platforms also offer more advanced features for requesters to maintain specialised groups of worker, e.g., based on demographic properties, worker skills, or requester-specific criteria.

The creation of microtasks can be supported by crowdsourcing platform providers both on a technical and conceptional level. On the technical level, a crowdsourcing platform can provide the infrastructure required to run a microtask. This can include resources like online storage for image upload, or software tools that can be used to generate surveys. On the conceptional level, crowdsourcing providers may provide best practices for microtask design, may recheck the requester's microtask design and correct common pitfalls, or may even create the microtask design for the employer.

Finally, crowdsourcing platforms provide means to process the microtasks submitted by the requesters. Here, the tasks might again be preprocessed by the platform, e.g., tasks may be replicated in order to enable quality control via majority voting. Then the microtasks are distributed to the workers. This can be either in an open call, i.e., the microtasks are publicly posted and any workers can decide to work on them, or a sophisticated worker selection can be performed, e.g., based on the workers' skills. After the workers complete the microtask, an optional post-processing of the results can be applied. This may include quality control or the aggregation of multiple submissions.

While all crowdsourcing platforms generally implement these three building blocks, different commercial providers put different emphases on each of them. In order to find an appropriate crowdsourcing platform for a specific research task, it needs to be clear which functions are required to successfully crowdsource the task. Consider a psychological study. Here, detailed knowledge about the demographic data of the participants can be of interest, i.e., detailed user profiles are required. In contrast, an image tagging task which is intended to create training data for a machine learning algorithm requires high quality results and consequently quality assurance mechanisms within the platform would be desirable.

3.2.0.2 Types of crowdsourcing platforms

As an intermediate step of identifying an appropriate platform for research tasks we will discuss three different types of crowdsourcing platforms: *Mediator crowdsourcing platforms*, *specialised crowdsourcing platforms*, and *platforms focusing on crowd provision* [22]. This coarse-grained categorisation can easily be applied to most existing crowdsourcing platforms and can be used for a first filtering of possible platforms. Figure 3.1 illustrates the types of crowdsourcing platforms and their interactions. In the following we briefly summarise the main aspects of the platform types and illustrate them with some commercial providers as described in [22].

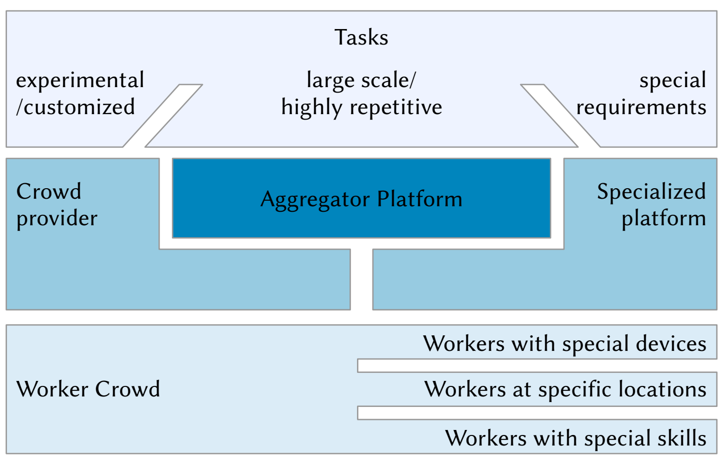

Fig. 3.1 Classification of crowdsourcing platforms.

Crowd providers are the most generic type of platform and mainly focus on building large-scale worker crowds. They provide means for accessing and managing the available workforce, e.g., filtering mechanisms, demographic information about the workers, and support for worker remuneration. Due to the direct access to the workers, these platforms allow for easily creating experimental tasks or building specialised enterprise solutions. However, due to this flexibility it is generally not possible for the platform operator to provide general purpose quality control mechanisms suitable for every use case. Platforms like AMT,[4] Microworkers,[5] RapidWorkers,[6] or ShortTask[7] are typical crowd providers. In a broader sense online Social Networks can be considered crowd providers. They can provide access to a large workforce but do not implement task routing or worker management systems.

Specialised crowdsourcing platforms maintain their own worker crowd and only focus either on a limited subset of workers (Crowdee[8] or Streetspotr[9]) or a specific type of microtask, e.g., Microtask[10] that mainly focuses on text digitalisation. Specialised crowdsourcing platforms like Microtask provide elaborate workflows for certain use cases. In this case the users of the platform have no influence on the actual microtask design, and only contribute to the data that will be processed. Platforms focusing on specialised workers, e.g., with specific devices or skills, allow a more flexible microtasks design which can be customised by the requester. However, there are often more restrictions on the microtask design than on crowd provider platforms.

Aggregator platforms focus on developing crowdsourcing-based solutions for large scale customers or general business cases. Similar to specialised crowdsourcing platforms, requesters using these platforms only need to submit the input data, while the actual microtask design is done by the platform. In contrast to specialised crowdsourcing platforms, aggregator platforms do not maintain their own crowd, but use the workers from crowd providers or the services from specialised platforms. Moreover, some aggregator platforms offer a self-service option where requesters can freely design their microtasks. As with crowd providers, no quality assurance mechanisms are offered here but the additional business layer between the requester and the worker is added, resulting in higher costs per microtask. Currently available aggregator platforms focus on business related microtasks, e.g., CrowdFlower[11] or CrowdSource[12] for content moderation or image tagging. Table 3.1 sum-

---

[4] Amazon, Inc. "Amazon Mechanical Turk". http://www.mturk.com (accessed April 2016).

[5] Weblabcenter, Inc. "Microworkers". https://microworkers.com (accessed April 2016).

[6] RapidWorkers. "RapidWorkers". http://rapidworkers.com/ (accessed April 2016).

[7] ShortTask. "ShortTask". http://www.shorttask.com/ (accessed April 2016).

[8] Crowdee, "Crowdee". https://www.crowdee.de (accessed April 2016).

[9] Streetspotr GmbH. "Streetspotr". https://streetspotr.com/ (accessed April 2016).

[10] Microtask Ltd. "Microtask". http://www.microtask.com/ (accessed April 2016).

[11] CrowdFlower Inc. "CrowdFlower". http://www.crowdflower.com (accessed April 2016.

[12] CrowdSource Solutions Inc. "CrowdSource". http://www.crowdsource.com (accessed April 2016).

marises the main characteristics of the introduced crowdsourcing platforms types.

Table 3.1 Crowdsourcing platform categories

	Crowd provider	Aggregator platform	Specialised crowdsourcing platform
Own worker pool	Yes	Yes	No
Costs per microtask	Low	High	Medium
Focus on specific microtask set	No	Yes	Yes
Predefined quality assurance mechanisms for specific microtasks	No	Yes	Yes
Unfiltered access to workers	Yes	No	No
Suitable for experimental tasks	Yes	Sometimes	Sometimes
Exemplary platform providers	AMT, Microworkers	CrowdFlower, CrowdSource	Microtask, TaskRabbit, Streetspotr

3.2.0.3 Applicability of crowdsourcing platform types for research

For most scientific use-cases *crowd providers* are the platform of choice. They allow direct, unfiltered access to the workers, enabling researchers to conduct tests on, for example, novel quality assurance mechanisms or incentive schemes, or conduct sociological or demographic studies. Moreover, the platforms usually allow requesters to create individual microtask interfaces on external servers that are required for experimental tasks or research on task design principles. However, running experimental tasks on *crowd provider* platforms usually requires higher conceptual efforts, e.g., because no quality assurance mechanisms are applied by the platform. Also the higher technical requirements cannot be neglected, as the microtask interface has to be provided by the requester, possibly along with the infrastructure for the workers to work on.

Sometimes there is a need to run experiments exclusively on specific user groups, e.g., crowd-sensing tasks or experiments about perceptual quality on smart devices. In this case, *specialised platforms* can be helpful, as some of them provide easy access to those groups. However, *specialised platforms* often offer predefined interfaces for the workers, thus it might be difficult to run experimental tasks on these platforms. Moreover, *specialised platforms* focusing on a specific task, e.g., transcription of handwriting, are only of value for the research community if exactly that task is required.

Aggregator platforms are most suitable for research tasks that are closely related to the main business focus of the platform, e.g., tagging of different image content. In this case *aggregator platforms* can support the scientists with means of quality control, interface design guidelines, or even provide the required infrastructure. If the task is not related to the platform's main business cases, *aggregator platforms* provide as little help as *crowd providers*. However, *aggregator platforms* sometimes apply filtering mechanisms to their user base that are not transparent to the requester. Thus, the crowdsourcing participants might be biased due to these mechanisms, while the researchers are not aware of this fact. Further, *aggregator platforms* are often more expensive then *crowd provider* as they represent an additional commercial layer between the requester and the workers.

3.2.0.4 Use of existing crowdsourcing platforms for research

Most existing crowdsourcing platforms can be assigned to one of the previously mentioned categories, which can serve as a first step towards finding the right platform for a scientific task. However, even if platforms belong to the same category, they can still differ in the supported types of tasks, demographics of their users [20], and their features for requesters and workers [48]. In particular, the platform access, the diversity of participants, the costs per microtask and for qualification tests, payment features, the performance to acquire testers, and the integration of the measurement software into the platform must be considered while selecting a platform for crowdsourcing scientific tasks.

AMT initially became popular for collecting research data, especially for US researchers. However, access to AMT has become more and more restricted both for requesters and workers in most countries, resulting in biased platform demographics. Due to the platforms payout policy,[13] the vast majority of workers are from India and the USA.[14] Requesters need to provide a U.S. billing address,[15] which also significantly limits access for non-US users. Therefore, recent work, e.g., by Vakhara et al. [48] and Peer et al. [39], tries to find and evaluate alternative platforms for crowdsourcing research, but finding an appropriate platform is difficult due to the high diversity of the platforms, their sheer number, and newly emerging enterprises.

One comercial crowdsoucing platform aimed at scientific tasks is Prolific Academic.[16] They provide extensive demographic information and support

---

[13] Amazon.com, Inc. "Worker Web Site FAQs". https://www.mturk.com/mturk/help?helpPage=worker\#how_paid (accessed July 2016).

[14] Panos Ipeirotis. "mTurk Tracker". http://demographics.mturk-tracker.com/\#/countries/all (accessed April 2016).

[15] Amazon, Inc. "Support for Requesters outside US on MTurk". https://requester.mturk.com/help/faq\#do_support_outside_us (accessed April 2016).

[16] Prolific Academic. "Prolific". http://prolific.ac/ (accessed April 2016).

the usage of well-known external survey tools like SurveyMonkey.[17] Prolific Academic is a young platform, so the sustainability of their business model is still unproven. Additionally, it needs to be determined if crowds exclusively working on scientific tasks, like surveys and subjective evaluations, will become highly biased. It has already been shown that even workers on AMT exhibit a growing non-naivety to typical research tasks [6].

The remainder of this chapter will shed light on some of the technical aspects that would significantly improve the usability of crowdsourcing platforms for use in research.

3.3 Proposed features to support academic research

The previous section outlined the capabilities of existing crowdsourcing platforms. This section examines the technological possibilities for enhancing such platforms to support their use in academic research.

3.3.1 User management

In this subsection we look at desirable features aimed at improving the crowdsourcing experience for both academic requesters and their workers. Problems with population sampling have been identified by various studies [7,15], hence greater access to relable user profiles is likely to reduce these issues. In many cases existing crowdsourcing platforms or third party add-ons have provided basic functionality, but more advanced and integrated features may allow requesters to target the most appropriate workers and so get better data. For the worker this leads to lower rejection rates for their work and makes it easier for them to find the best paying microtasks.

3.3.1.1 Worker profiling

Current crowdsourcing platforms do include the ability to find limited user profile information, usually about the abilities of the workers in relation to microtasks performed. However, information about basic demographics such as age, sex, location or education level is typically not accessible. Either this information is not stored by the platform or is hidden from view. Academic requesters often need access to this information, either to restrict a study to a particular subset of the population, or to examine differences in results

[17] SurveyMonkey Inc. “SurveyMonkey”. https://surveymonkey.com (accessed April 2016).

from various demographics. For instance, a study might look into the differences in understanding of computer security by different age groups. However, it should be noted that providing increased information about workers is not without risk to their privacy. As the number of data points about a worker increases, the potential for requesters to be able to successfully use de-anonymisation techniques to find their identity also increases [37].

As a result, demographics should be released to requesters with caution. Perhaps a more in-depth relationship between requester and platform might allow access to such information. One could imagine the requirement for evidence of ethics approval, plus a demonstrated commitment to data and worker confidentiality as a subset of the requirements for such certification. Similarly, those workers who could verify their profile and were happy for it to be released to certified requesters could access more interesting and well paid work. Characteristics such as physical attributes or medical conditions may be also be included in such demographics, but this sensitive data introduces further legal and moral issues.

Abilities and characteristics can be measured by computerised tests. There are numerous tests for English comprehension, colour blindness and other features of vision [16] (see Chapter 4). There are batteries of tests for various cognitive abilities including spatial, intelligence and memory [3], including the well known, but controversial IQ tests. Personality may also be measured, for instance, using BFI tests [2]. Workers declaring background knowledge and domain expertise in a particular area may also be of interest, and can be verified online. For example, a requester may be interested in examining the abilities of computer programmers when presented with particular problems. However, testing online leads to concerns about cheating and also tuning to perceived biases from requesters. The latter is particularly an issue in personality tests. Cheating, where the worker gains help from others, can be mitigated to some extent by the platform initiating the tests randomly, rather than the worker starting the test at their convenience when they are prepared and have resources to hand. The issues with tuning test results are more difficult to counter. A major impact on avoiding problems of this sort is to increase the trust within the worker, requester and platform relationship. One factor here is the number of microtasks that have an upper bound in tests, rather than a lower bound. For instance, researchers may be interested in evaluating computer interfaces for those with low cognitive ability. When there are sufficient microtasks that have an upper bound on test results, cheating on such tests is less of a problem as there is no obvious advantage to having a better performance.

3.3.1.2 Worker hardware details

Another dimension of user management that is subtly different from worker profiling involves understanding the technology that the worker is using to perform any particular microtask on. Gaining access to detailed information about screen size and resolution, input mechanism, operating system, internet connection speed—even whether the participant is having to scroll within the page to answer the question—would be useful information in many studies. This sort of information could allow a requester to restrict studies to those with a minimum screen size, or only to those accessing the study via a mobile device.

3.3.1.3 Reputation management

The reputation workers have on crowdsourcing systems is a powerful driver of behaviour as it encourages workers to be more accurate and reliable. Workers concerned about maintaining their reputation are more likely to accurately state their abilities and skills. The main motivation for workers to maintain a good reputation is that they can get better paid jobs, see Section 3.3.2 on payments and motivation. At present worker reputation is restricted to single sites and revolves around measuring job performance accuracy and acceptance/rejection rates from requesters. Requesters also have a reputation and those with good standing attract more and better workers. Requester reputation is typically measured by workers, and can be across a number of factors, such as promptness of payment and generosity. Worker reputation is managed by the crowdsourcing site, however, access to requester reputation is via third-party sites, such as Turkopticon [23].

Current tools for reviewing completed microtasks on crowdsourcing systems are seriously limited. AMT restricts reviewing of microtasks (HITs) to only acceptance or rejection of work. Microworkers, allows the requester to ask workers to revise a microtask instead of just rejecting it. However, a more fine-grained approach would allow the requester to give feedback on performance without resorting to the 'binary' option of refusing payment. The quality of feedback from requesters could then be part of the profile required for particular jobs.

There is a strong case to be made for crowdsourcing platforms to directly manage information about the reputation of requesters. This is typically not a feature of current platforms. As noted above, requester reputation information is usually only available via third party sources. The current situation might be considered problematic for requesters as it has the danger of incomplete information and lack of redress and so there is the potential for malicious and inaccurate information about requesters to be circulated. Adding requester reputation information onto current platforms would mitigate against these issues, making this information more reliable. The consequence for workers is that they are provided with more accurate information about the requester when choosing jobs.

The impact of reputation might be even stronger if it could be transferred across different platforms. This would avoid platform lock-in and foster more open and flexible digital labour markets. Technically this would be feasible through a server-client approach, where reputation is managed centrally, and crowdsourcing platforms communicate with a reputation server. An alternative solution would be to maintain a peer-to-peer architecture, allowing a more flexible approach to platforms leaving and entering the network. In either case, defined web service standards for distributing encrypted reputation information are needed. However, since most crowdsourcing platforms are commercial entities, they would prefer to keep requesters and workers on their system, and there is very little motivation for them to provide this sort of functionality.

3.3.1.4 Requester-worker communication

As noted elsewhere the requester-worker relationship is unbalanced, with workers having little recourse when payment is refused. The main communication channel is typically email, which removes anonymity and usually does not allow communication before or during a microtask. Channels to enable more immediate, confidential and anonymous communication between workers and requesters are easily within technical grasp: chat systems and message boards are now familiar through prevalence in social media sites. These could integrated into crowdsourcing platforms. Workers would get a mechanism for getting clearer, interactive instruction and a more controlled system for raising concerns about payment. The advantage for requesters is better communication about complex and time-consuming microtasks. As academic studies are often some of the more sophisticated microtasks, research using crowdsourcing stands to benefit from improved communication. There is however a challenge of requester availability for communication, particularly across time zones.

3.3.2 Payments and motivation

In many contexts workers expect some kind of reward for their participation in collaborative activities and experiments. For some microtasks this can be intrinsic, such as collaboratively building an encyclopaedia, for others this is direct financial reward.

3.3.2.1 International payments

Most commercial crowdsourcing platforms have a primary currency. For Germany-based Crowdee it is the Euro; for British platform Prolific, it is Pounds Sterling; and for AMT, the U.S. dollar. This means that for a proportion of the user-base payments must cross financial borders. Exactly how this is processed depends on the chosen platform, the currencies available to the requester, and the supported deposit facilities of the worker. AMT, for example, supports direct deposit payments to U.S. bank accounts in U.S. dollars and Indian bank accounts in Indian Rupees only. All other workers are issued payment only as Amazon.com gift cards.[18]

3.3.2.2 Alternative payments

Going forward, new payment options may help sustain and grow the crowdsourcing labour market. It may already be possible to make payments outside of existing frameworks. For example, international payment processors like PayPal offer an alternative when the platform does not directly support payments or where payments are difficult and payment costs are prohibitive. More novel payment methods, such as Bitcoin may also be used to support more anonymous payments to workers. These alternative payments may raise additional issues with regard to circumventing commission charged by platforms, obligations for transaction traceability within the requesting organisation, or with local laws.

3.3.2.3 Legal concerns

A particular concern for platforms is their legal liability for tax and money laundering. Crowd providers may try to minimise any potential involvement in an employer-employee relationship and any potential tax liability or labour responsibilities arising from it. Some platforms, like AMT, have strict sign up requirements for workers and requesters alike. AMT requests personal information, including tax reference numbers for requesters who deposit money for the platform as well as those receiving payments.[19] AMT gathers this

---

[18] Amazon.com, Inc. “Worker Web Site FAQs”. https://www.mturk.com/mturk/help?helpPage=worker\#how_paid (accessed July 2016).

[19] Amazon.com, Inc. “Requirements for Purchasing Prepaid HITs”. https://requester.mturk.com/mturk/amazonpaymentsacctreqmts (accessed April 2016).

information to support their legal reporting obligations with regard to both the U.S. Patriot Act and Internal Revenue Service (IRS) regulations.[20]

Another consideration for requesters is the employment status of workers. In many cases workers are considered independent contractors, hired by the requester. While this approach may limit legal liability for the platform provider and requester alike some jurisdictions may consider regular, repeat workers of a given requester to be eligible for additional rights and benefits [12] such as healthcare or pension contributions. In these cases it may be important to restrict repeat patronage of a given worker to limit unintentional additional liabilities. Many platforms provide a consistent worker ID and it can be recorded, along with microtask durations, to allow zealous workers to be excluded from future microtasks if needed.

While platforms may supply some documentation and support for requesters and workers, it is important to consider any local implications for cross-border payments and any inferred employment relationship that payment may create. Exact liabilities may not be immediately obvious and should be thoroughly investigated before carrying out crowd-work, especially on an ongoing basis.

3.3.2.4 Non-monetary rewards

Workers may also be encouraged to participate by offering non-financial incentives. In the case of tasks such as the usability evaluation of a software product, workers may consider early access to unreleased software as a sufficient incentive to participate. Large collaborative projects like Wikipedia provide a product directly to the user base and encourage a collective ownership [9]. Similarly the popular "citizen science" project *Galaxy Zoo* and later the *Zooniverse*, depends on a variety of intrinsic motivations among their participants to support the categorisation process. Here, participants are engaged by appealing to their enjoyment of astronomy, learning and discovery, and their willingness to contribute to scientific research [40, 41].

Participants can also be rewarded by providing them with their own processed data. Seeing how they compare to other workers is a core concept of "gamification". By improving the enjoyment and competitiveness, workers can be encouraged to better engage with the microtask. This approach gives workers a target or goal that they wish to meet to highlight their own competence and can lead to a higher efficiency and improved quality [10]. Also, workers may become engaged with the scientific process and be motivated by seeing their contribution, for example in extreme cases workers have become

[20] Amazon.com, Inc. "IRS Reporting Regulations on Third-Party Payment Transactions For Personal or Business Account Holders". https://payments.amazon.com/help/200831230 (accessed July 2016).

so engaged in the research outcome as to warrant authorship of published work [43].

3.3.3 Ethics

Here we summarise issues connecting ethics with technology when using crowdsourcing tools. We note that Chapter 2 has a more detailed and general discussion on ethics in crowdsourcing. When such technology is used to support academic research one needs to consider ethical questions from different perspectives, related to the two roles of the participating humans (workers):

- Objective – Workers are active, conscious participants of the research effort, providing their expertise to help obtain, process or interpret scientific data. Examples include protein folding,[21] space exploration [41].
- Subjective – Workers are subjects of the research, where the crowdsourcing platform acts as the environment for experiment execution, during which the workers are observed interacting with the platform, microtasks and other workers. Examples include evaluation of working patterns [30], evaluation of monetary incentives [33]

3.3.3.1 Objective participation

Workers participating in crowdsourced experiments need to be clearly informed about the conditions of their participation. Usually this implies explicitly stating the participation conditions beforehand (description of requested contribution, time constrains, rewards), presenting ethics approval for the experiment from a trusted organisation and requesting the participant read and accept this, and stipulating how sensitive data will be handled.

Apart from legal reasons, being informed about the precise participation conditions and the effects the participant's contribution may have on the overall outcome is important because many crowdsourced research efforts are based on volunteering, and it has been shown [14,19,31] that the expectation of the positive contribution to the science is the principal motivational factor in this case. At the same time, not being clear on the participating conditions demotivates many participants who fear that providing subpar contributions will harm the overall effort, which often leads to high attrition rates. Regardless of the fact that many workers are willing to contribute voluntarily to various scientific efforts, the experiment organiser needs to be aware that the experiment they run still represents an exploitation of, otherwise expensive, cognitive labour. This is why it is important to compensate for the missing

[21] FoldIt. "Solve puzzles for science". https://fold.it/ (accessed April 2016).

or symbolic monetary rewards by introducing a set of psychological incentives acting on the intrinsic motivation of the participants and helping them achieve a sense of self-fulfillment. An informative case study can be found as part of the Smart Society project.[22]

Storage of sensitive data must be considered from both legal and technical perspectives. Both can have direct ethical implications. The information contained in the stored data should be reduced to the minimum needed for successful functioning of the platform and execution of the experiment. Techniques such as data anonymisation and semantic obfuscation [11, 17] can be used to reduce the exploitability potential of the stored data. The simplest examples include storing age range instead of concrete age (birth date), and storing geographical area instead of concrete address. Even when appropriate care is taken to assure the protection of sensitive user data, one should consider third-party services as well. Consider, for example, a crowdsourced study where participants are asked to provide personal anonymised health data and are rewarded with monetary rewards. Even when the experiment organisers act in best faith and follow all precautions for keeping the health data anonymised, poor management of payment data can allow matching the two datasets and ultimately breaching the promised data policy. It is therefore advisable to choose a crowdsourcing platform which can guarantee a safe and separate handling of payments, or delegate the payment management to a trusted third party (cf. Section 3.3.2). The choice of the payment processor and the payment data retention policy should also be clearly stated in the consent form, together with the country-specific conditions which may apply.

3.3.3.2 Subjective participation

Crowdsourced experiments where the workers are subjects of the study are typical in social sciences and experimental economics. They generally involve use of general-purpose crowdsourcing platforms where the experiment setup is obtained through a combination of a specific microtask design, worker selection procedure and the set of incentives (rewards). Selected workers are commonly divided into experimental and control groups, and are usually not aware that they are taking part in an experiment, as this might otherwise yield skewed results. During such experiments, the microtasks given to the workers may (purposefully or not) exhibit properties that will cause certain behavioural responses to be more accentuated than for an average microtask, e.g., fatigue, drop of concentration, sense of insecurity, frustration, competitiveness. Since many people working as crowdworkers receive a significant amount of income [32] this aspect becomes increasingly important with the

[22] SmartSociety Consortium. "Deliverable 5.3 - Specification of advanced incentive design and decision-assisting algorithms for CAS" http://www.smart-society-project.eu/publications/deliverables/D_5_3 (accessed July 2016).

potential to affect daily lives.[23] If an experiment is expected to cause the described effects, the experiment setup should include distraction and leisure tasks or incentives. For example, a common strategy for image tagging microtasks is to occasionally offer interesting and funny pictures to the crowd. Similarly, in Galaxy Zoo project, participants are occasionally shown easy pictures to boost their self-confidence, or even sent personalised motivational messages.[24]

The aforementioned issues are just a part of a wider debate on worker rights that is currently raising much interest in the research and the worker community (seethe Fair Crowd Work website[25] for a compilation of relevant topics). Currently, the working conditions are determined solely by the crowdsourcing platforms and the requesters. This means that crowdworkers are often treated as isolated individuals and harnessed as 'human subroutines'. This has in turn lead to self-organisation of crowdworkers using alternative, independent forums or platforms, such as Turkopticon [23]. This has direct implications for requesters as well, since the requester's reputation among the worker population can determine which workers will accept the microtask and under which conditions, potentially affecting the outcomes of the experiment. Therefore, fair microtask rewards and execution conditions become important factors to consider when designing a crowdsourced experiment.

At the same time, these worker self-organisation platforms are also allowing the workers to share hints and advice on gaming a particular requester to maximise their rewards for the smallest amount of effort. While data quality control is necessary in most crowdsourcing efforts since part of the worker population will always be producing subpar results (see [9]), integrating robust mechanisms for quality control and incentive mechanisms becomes even more important for crowdsourced experiments as they usually offer microtask compensations that are higher than the average, thus attracting attention of malicious users and prompting their exploitative actions.

Apart from providing a means to collectively defend worker rights, the self-organisation platforms are also a tool for today's crowdworkers to socialise and establish informal communities. While native support for socialisation is an expected [27] property of future crowdsourcing platforms, for an experiment designer this will pose yet another important trade-off to consider, particularly during longitudinal studies: It has been shown [30, 45] that socialisation and communication among workers can significantly affect task outcomes and thus the experiment itself, e.g., by possibly 'contaminating'

[23] Harris, Mark. "Amazon's Mechanical Turk workers protest: 'I am a human being, not an algorithm"'. http://www.theguardian.com/technology/2014/dec/03/amazon-mechanical-turk-workers-protest-jeff-bezos (accessed April 2016).

[24] SmartSociety Consortium. "Deliverable 5.3 - Specification of advanced incentive design and decision-assisting algorithms for CAS" http://www.smart-society-project.eu/publications/deliverables/D_5_3 (accessed July 2016).

[25] Fair Crowd Work. "Fair Crowd Work". http://prolific.ac/ (accessed April 2016).

the control group. The key thing to consider here is finding a fair way to maintain the experimental setup, while not isolating the workers. We are not aware of any standards or widely agreed-upon conventions regulating the worker organisation and socialisation; each crowdsourcing platform is free to decide if and how to implement support for such functionalities. Therefore, the experiment designer must consider this on a case-by-case basis.

3.3.4 Additional instrumentation

Basic reporting of results is a staple of crowdsourcing platforms but is often limited to a simple key-value store for each question. Additional instrumentation can be beneficial to better understand user engagement with microtasks, especially in experimental settings.

Platforms vary in their ability to support monitoring of worker behaviour and their devices. For example, web-based platforms such as AMT will not be able to provide direct access to hardware sensors [42]. For web-based platforms the availability of technologies and abstractions supported by the browser including JavaScript and HTML5 will impact study design and collected data. Device-focused studies can offer much more comprehensive data collection opportunities and provide richer context awareness [13]. However, app-based platforms may require more extensive programming and may narrow the diversity of workers for a microtask or result in an unintentional selection bias.

3.3.4.1 Behaviour monitoring

Additional behavioural data can contextualise existing findings and offer new avenues for research into user behaviour in crowdsourced environments. By capitalising on existing inputs in new ways, a richer understanding of worker behaviour can be discerned. Recording additional user information can also provide validation and verification of the primary data. As workers may employ techniques to minimise the time spent working on microtasks—such as automation or more complex group activities—additional understanding of user activity is vital to gathering high quality data [8].

Keyboard and mouse

As the primary input devices for non-touchscreen devices, the keyboard and mouse can provide significant insights into user interactions [38]. Recording keyboard and mouse events is possible, even in web-based platforms such as

AMT. For basic interactions—such as to identify the order in which questions were attempted or whether the user left the microtask—timestamped actions, such as `focus` and `blur` (unfocus) events, can be recorded.[26] This allows for detection of which items are selected and deselected and can also be used to track when the web page showing the microtask is in the foreground. This indicates whether a user is fully engaged with a microtask and can aid in identifying multitasking or the use of external resources. For more complete analysis of user activity full keyboard and mouse interactions can be recorded. Events including `keypress`, `mousemove`, and `click` are fired when users engage with the microtask. By recording these interactions a comprehensive picture of user activity can be built and analysed, or even played back [4] for example in evaluating the evolution of a users design [26]. Additionally, these user interactions can be correlated with accuracy and, going forward, be used as a potential indicator of the quality of a worker's efforts [21,25,36].

Audio and video

Another commonly available input is audio and video. Audio recording can be used to capture user thoughts and support think-aloud protocol experiments, while video offers a variety of user engagement opportunities such as eye-tracking [28], emotion detection [34], and augmented reality [46]. For web-based studies the Adobe Flash plugin provides a widely deployed platform that can be used to allow audio and video inputs to be captured [35]. Similarly, the emerging HTML5 WebRTC API provides plugin-free support for capturing audio and video [28]. This data can be uploaded to a server either in real-time or after microtask completion depending on the experiment needs. However, this type of monitoring of user interactions in an otherwise uncontrolled environment may raise privacy concerns for users and researchers alike [1].

Combining techniques

Where techniques such as audio recording are problematic for otherwise anonymous remote interaction, surveying may provide an alternative. Surveying the user on their thoughts both about the microtask and how they chose to carry it out can provide richer qualitative information that may otherwise be missed in these interactions. Simply asking the user to indicate how long they have spent on a task, noting their absences or engagement can provide an increased insight over a purely technological approach to measuring engagement.

[26] World Wide Web Consortium (W3C). "UI Events Specification". https://www.w3.org/TR/uievents/ (accessed February 2016).

Some services such as *Upwork* (previously known as *oDesk*) use a combination of monitoring and surveys. Their software client both asks users to record time worked (which is used for billing purposes) and allows requesters to inspect details of key presses, mouse movements, and periodic screen shots [5]. While the Upwork model more closely mirrors a typical employer-employee relationship, the pseudo-anonymised nature of many crowdsourcing platforms limits the acceptance for this type of monitoring. However, as the prevalence of both technological support and user acceptance for audio-visual recording grows, it may become practical to reintegrate these methods into crowdsourced-based research.

3.3.4.2 Emerging opportunities

Combining existing sensor technology, emerging browser and device support, and new algorithms, further advances in user monitoring can be achieved. Once seemingly limited to keyboard-based desktop-bound tasks, crowdsourcing has become far more mobile, and with a much broader input modality [44].

Mobile devices

Consumer mobile devices commonly include a multitude of sensors including location sensors and movement sensors. In web-based environments, these sensors are abstracted and supported by the Geolocation API[27] and `devicemotion` events.[28] Geolocation can support "in the wild" crowdsourcing of data, such as generating location-based datasets. Additionally, device motion offers opportunities for unique device interaction techniques and can aid in recognising user activity [18]. Newer devices offer additional dedicated sensors such as pedometers and heart rate monitors. As these devices become more common and their interfaces are standardised, additional data collection opportunities will emerge.

Eye tracking and biometrics

Understanding what engages users can provide important pointers for improving microtask design and research outcomes [24]. Eye tracking offers an improved measure of what parts of a microtask attract the most attention compared to mouse tracking [29]. By tapping into the nearly ubiqui-

[27] World Wide Web Consortium (W3C). "Geolocation API Specification". https://www.w3.org/TR/geolocation-API/ (accessed March 2016).

[28] World Wide Web Consortium (W3C). "DeviceOrientation Event Specification". https://www.w3.org/TR/orientation-event/ (accessed February 2016).

tous webcam, identifying salient features of on-screen images can already be achieved [50]. Video can also lend itself to biometric monitoring—offline video processing to highlight seemingly imperceptible changes such as breathing and heart rate has been demonstrated [49]. Using such processing in real time has the potential to offer biometric data from already deployed sensors.

3.3.5 Supporting different study designs

In considering the question of whether crowdsourcing technology can support academic research, it is necessary to discuss the needs of academic research studies beyond traditional surveys offered as microtasks through crowdsourcing services like AMT. Research studies can vary greatly in their design, however there are some common considerations that affect many of these study designs.

Most researchers are not computer programmers, but they follow certain processes in order to conduct rigorous academic research. This means if they are to utilise crowdsourcing platforms for conducting studies, these platforms will need to provide out-of-the-box support for some common types of studies. This section talks about the requirements for different study designs as they regard to potentially conducting academic research through crowdsourcing platforms.

While most of the existing crowdsourcing platforms discussed earlier do not directly support academic research, there are several that cover a subset of the desired features. Qualtrics[29] provides online software specifically for running customer experience surveys. As mentioned earlier, Prolific[30] is a crowdsourcing platform specifically designed for conducting academic studies. Some of Prolific's features are: high-quality participants, flexible prescreening, support for longitudinal research, bonus payments based on quality.

3.3.5.1 Focus

One of the issues with current crowdsourcing platforms is the potential lack of focus of the workers during data collection. Traditional studies often compare times taken to achieve a task, or gauge reaction to one stimuli after viewing another. In-person studies can carefully control for variables such as external stimulus, distractions and time between stimuli. For example, many studies will put a participant in a quiet, empty room free of distractions. However these factors are almost impossible to control when using crowd-

[29] Qualtrics LLC. "Qualtrics". http://qualtrics.com/ (accessed April 2016).
[30] Prolific Academic. "Prolific". http://prolific.ac/ (accessed April 2016).

sourcing approaches where people are completing projects in a variety of locations surrounded by potential distractions. It is possible to work around this problem by designing studies around the constraints of the platform, but in general this is a major hurdle to academic research being conducted on crowdsourcing platforms.

3.3.5.2 Interactivity

Many Human-Computer Interaction (HCI) experiments involve participants interacting with software. Crowdsourcing platforms are appealing for such research from the perspective of attracting and managing participants, and for the number of participants they potentially provide. However this sort of usability research often tests software with novel user interfaces. Earlier in the chapter we discussed how crowdsourcing platforms are predominantly web-based, and are limited in the customisability of web pages presented as microtasks for workers. The diverse background, locations and computer capabilities of crowdsourcing workers means that any software being studied is almost certainly required to be web-based. Luckily, increasing numbers of software applications (especially research applications) are being built using web technologies. This increases the prospect that these applications could be presented and studied via crowdsourcing platforms. However some of these web applications can still be quite demanding in their resource needs (e.g., processor speed, bandwidth, persistent connection to a remote server) which could be a problem for workers without fast computers or reliable connections. This information about worker's hardware capabilities could be evaluated by the crowdsourcing platform and used for participant selection.

HCI experiments often use specific study designs in order to collect data they want. For example, in some experiments all participant interaction with the software will be recorded. Historically this has been done with a video camera pointed at a screen, with screen recording software, or with the application recording the individual interaction events. The purpose of this data collection is to see what the participant did. In the case of crowdsourced workers, only the last of these options is really feasible. It is definitely possible to instrument a web application in this way, but it is not trivial and requires a large amount of additional work on the part of the developer or researcher.

Another approach used for in-person usability studies is to have a researcher observe the actions made by participants, and for the researcher to take notes of interesting events and discuss these with the participant once the experimental task is complete. Remote workers in different time zones mean such observation and discussion—if possible at all—would need to involve the researcher viewing the experiment after the fact and contacting the worker to obtain feedback. This requires the cooperation of the worker to provide this follow-up feedback at a time when they may no longer recall their actions or the motivation behind them.

Ultimately such studies aim to determine places where participants do or do not understand the intentions of the interface, and therefore where they can or cannot use it effectively. This requires a significant understanding of the participants' reasoning while performing actions. Another approach for this is to use a think-aloud protocol and get the participant to (try to) verbalise their thinking behind the actions they are performing. This is a very effective tool for usability evaluations but generally some amount of questioning and prompting from a study facilitator is required to get the necessary data (i.e., keep them thinking aloud). Even assuming that crowd workers are set up to record and transmit back audio, this prompting is not something that can be easily duplicated if experiments were being conducted in a crowdsourced setting.

3.3.5.3 Collaboration

An issue with existing crowdsourcing platforms potentially being used in research is the lack of support for collaboration. Many research experiments involve two or more participants working on a task simultaneously, or collaborating to reach a shared goal. To support this kind of research, crowdsourcing platforms would need to provide better support for collaboration. This is not a technological impossibility—an internet-connected software environment for running studies could obviously be extended to support the communication required to support collaboration. However this would likely require crowdsourcing platforms to move to a model where the requester (researcher) can be more directly involved in the data collection, i.e., they can interact with the workers in some way while the study is in progress in order to facilitate collaboration or discussion. This is in contrast to the current model where completion of microtasks produces data which is then processed by the requester at a later date. This has also further implications for the quality control mechanisms, as currently most mechanisms use the assumption that the submissions of the workers are independent of each other.

A common qualitative data collection technique is to conduct *focus groups*. Focus groups involve a group of participants being shown or told about something, and then providing their opinions and thoughts on the thing in question via a group discussion. Focus groups require the researcher to facilitate the discussion with prompting questions. It is a very effective technique, but one that is hard to translate to the crowdsourced environment, both because of requiring worker-to-worker communication but also the involvement of a facilitator.

In both these case we see that a research-oriented crowdsourcing platform would almost certainly require come capability for researchers to communicate or interact with workers while they are completing microtasks.

3.3.5.4 Randomisation, group assignment

Some common study designs are *between-groups* experiments or *repeated measures*. Between-groups experiments get similar groups of participants to do the same task while keeping all but one variable the same. The groups can then be compared to determine the effect of the variable on the task. It is obviously important to control the number of differences between groups that could be a confounding variable. Such experiments also require participants to be randomised between groups and for groups to be balanced. These last two needs would be easy to address in a crowdsourcing-based experiment, but controlling for confounding variables is difficult to do when there is no direct control over the environment in which the worker does the study, and whether they have access to external resources.

Repeated measures experiments use the same participants and get them to do all tasks under the different sets of conditions. In this case it is important to get participants to perform tasks in a randomised other. Possible confounding factors could be introduced by participants conducting the study over a longer period in multiple sittings, or due to the researcher being unable to control the environment in which the participant completes the task.

3.3.5.5 Longer studies

Many research studies repeatedly collect data over long periods of time from the same participants. Such studies are known as longitudinal studies. The longest longitudinal studies are over 75 years, e.g., [47]. These kinds of studies are difficult to conduct and tend to lose many participants over time.

Such studies should not be any harder to conduct using online crowdsourcing platforms. In fact it might be easier in this environment since researchers could begin with a larger pool of participants, an online system can more easily remind or prompt people to participate, and it may be easier to keep track of people via email accounts than postal addresses (which are likely to be more transient).

Crowdsourcing platforms that were to support longitudinal research would need capabilities for repeating studies with the same participants, a targeted notification system, and support for incremental payments with a possible bonus for completing entire term of the experiment.

3.3.5.6 Participant selection

Scientific studies typically have some requirements in terms of selecting participants. Even if they do not select participants based on specific criteria it is usually necessary to report on the characteristics of the participants—i.e., their ages, gender, background, or any other attributes that could seen to affect the results. One such characteristic is familiarity—there might be a requirement that participants have not participated in a similar study before, or that they do not have any familiarity with the thing being tested. Current crowdsourcing platforms provide only minimal details of workers to the requester. It would be easy for crowdsourcing platforms to store additional details of their workers. This is also useful information for the platforms to have, since it is effectively information on the demographics of their workers.

Additionally, many research studies need the ability to automatically assign participants to different conditions (i.e., different participant groups who are given different tools or stimuli during the experiment). For example, a study might require different groups of participants to do different tasks. The work of handling this assignment to conditions and of randomising the experiment itself would usually be done by the researcher, but would require automation in crowdsourcing platform setting. For this, the platform would be required to understand details of the experiment, such as how participants are assigned to conditions, and how the experiment is structured for these groups, so that this information could be automatically applied when workers undertake the experimental microtask.

While research experiments may sometimes utilise a very small number of participants, it is common for these people to have specialised skills, e.g., for them to be subject experts in a particular domain. For example, a study may seek the opinion of people familiar with perception, visual algorithms, or interaction techniques. As noted in Subsection 3.3.1, a crowdsourcing platform could allow a worker to specify such expertise, but it may be necessary to have a mechanism for verifying such information. Also, subject matter experts may want to provide qualitative feedback on designs. Such feedback would traditionally be free-form, comprising of verbal feedback, written notes, annotation of paper designs, or gesturing. An online form can certainly be used to collect textual comments, but in order to get useful feedback of the same quality as in-person studies it might be necessary for research crowdsourcing platforms to provide a richer means of providing feedback. Some possibilities would be allowing video responses or web-based annotation of diagrams.

3.3.5.7 Activity-tracking studies

Many health, fitness, or product related studies get participants to record information about their daily activities, such as food intake, exercise activity, or purchasing decisions. The traditional method for conducting these studies involves participants keeping a journal of activities and submitting this to the researchers at regular intervals.

Such manual journalling is not ideal since participants may forget to enter some data, they may enter incorrect data (accidentally or by choice), or may make errors during data entry. Online systems, including a web-based crowdsourcing approach, have the benefit of being able to prompt or remind the participant to enter their data (especially when they are using a mobile browser). They can also validate data to check that, for instance, specified data is within a particular range or is close to expected values. Additionally, many classes of errors can be avoided because sensors on computers or mobile devices can be used to check values that a participant would otherwise have to check and enter manually. As discussed in Subsection 3.3.4 the rise in mobile computing means that information from a wide variety sensors could considered when designing studies. Some examples of such information are date, time, physical location and heart rate. Another benefit is that such a system can provide immediate feedback or advice to the participant, in additional to tradition participation payments discussed in Subsection 3.3.2. The requirements here are for research crowdsourcing platforms to be able to use device capabilities to check some values and to be able to validate other data that is entered and provide the participant with immediate feedback when it is not valid.

A recent example of collection of study data via mobile devices is Apple's ResearchKit. Introduced in mid 2015, ResearchKit is an iOS framework that developers can use to build apps for conducting scientific research via mobile apps. It allows participants to use their device for collection of study data, e.g., using the "accelerometer, microphone, gyroscope and GPS sensors in iPhone to gain insight into a patient's gait, motor impairment, fitness, speech and memory".[31] ResearchKit allows access to this data in a controlled manner that is clear to the participant. Collection of data via a sensor-rich mobile app has further benefits such as the fact that participants always have the device (and therefore the app) with them and that such apps can communicate with connected devices to collect data via additional sensors, such as a heart-rate monitor on a watch or fitness band.

3.4 Conclusions

In this chapter we examined how various platforms, technologies, and techniques can support crowdsourcing in an academic context. We first discussed the capabilities of existing public crowdsourcing platforms and outlined the

[31] Apple Inc. "ResearchKit". http://www.apple.com/researchkit/ (accessed June, 2016).

types of features they provide to requesters. We then discussed possible feature additions or enhancements that would benefit academic studies conducted via these platforms. The proposed features fall into the broad categories of user management, payments and motivation, ethics, additional instrumentation, and supporting different study designs.

Finally, we considered the advantages and disadvantages of crowdsourcing some broad classes of study design, including between-groups, repeated measures, and longitudinal studies. We discussed the particular needs of research-related microtasks and how some of these could also enhance or benefit existing (non-research) microtasks conducted on these platforms. Some of features we proposed included, richer demographic information for workers, better reputation tracking or certification to gauge worker quality, support for varied forms of payment, better microtask monitoring and communication channels between workers and requesters, and platforms support for study designs and enforcement of study procedures.

We suggest there are many relevant features that could be easily added to crowdsourcing platforms that would greatly increase their appeal to researchers. Many of these features are straightforward to implement and would benefit existing workers and requesters in additional to potential research users. While we recognise there are still significant hurdles to the wide adoption of crowdsourcing within academia, there are many easy steps that crowdsourcing platforms can take to increase their usefulness to such domains.

Acknowledgements

The genesis and planning of this chapter took place at the Dagstuhl Seminar #15481, "Evaluation in the Crowd: Crowdsourcing and Human-Centred Experiments" held in November 2015. Jason Jacques was supported by a studentship from the Engineering and Physical Sciences Research Council. Ognjen Scekic was supported by the EU FP7 SmartSociety project under grant #600854. Michael Wybrow was supported by the Australian Research Council Discovery Project grant DP140100077. This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG) under Grants HO4770/2-1 and TR257/38-1. The authors alone are responsible for the content.

References

1. Adams, A.A., Ferryman, J.M.: The future of video analytics for surveillance and its ethical implications. Security Journal 28(3), 272–289 (Jul 2015)
2. Barrick, M.R., Mount, M.K.: The big five personality dimensions and job performance: a meta-analysis. Personnel psychology 44(1), 1–26 (1991)
3. Bertua, C., Anderson, N., Salgado, J.F.: The predictive validity of cognitive ability tests: A UK meta-analysis. Journal of Occupational and Organizational Psy-

chology 78(3), 387–409 (2005)
4. Breslav, S., Khan, A., Hornbæk, K.: Mimic: Visual analytics of online micro-interactions. In: Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces. pp. 245–252. AVI '14, ACM, New York, NY, USA (2014)
5. Caraway, B.: Online labour markets: an inquiry into oDesk providers. Work Organisation, Labour & Globalisation 4(2), 111–125 (2010)
6. Chandler, J., Mueller, P., Paolacci, G.: Nonnaïveté among amazon mechanical turk workers: Consequences and solutions for behavioral researchers. Behavior research methods 46(1) (Mar 2014)
7. Crump, M.J., McDonnell, J.V., Gureckis, T.M.: Evaluating amazon's mechanical turk as a tool for experimental behavioral research. PloS one 8(3), e57410 (2013)
8. Difallah, D.E., Demartini, G., Cudré-Mauroux, P.: Mechanical Cheat: Spamming schemes and adversarial techniques on crowdsourcing platforms. In: CrowdSearch. pp. 26–30 (2012)
9. Doan, A., Ramakrishnan, R., Halevy, A.Y.: Crowdsourcing systems on the World-Wide Web. Commun. ACM 54(4), 86–96 (Apr 2011)
10. Eickhoff, C., Harris, C.G., de Vries, A.P., Srinivasan, P.: Quality through flow and immersion: Gamifying crowdsourced relevance assessments. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 871–880. SIGIR '12, ACM, New York, NY, USA (2012)
11. Elkhodr, M., Shahrestani, S., Cheung, H.: A semantic obfuscation technique for the internet of things. In: 2014 IEEE International Conference on Communications Workshops (ICC). pp. 448–453 (June 2014)
12. Felstiner, A.: Working the crowd: Employment and labor law in the crowdsourcing industry. Berkeley Journal of Employment and Labor Law 32(1), 143 (Aug 2011)
13. Ferreira, D., Kostakos, V., Dey, A.K.: AWARE: Mobile context instrumentation framework. Frontiers in ICT 2 (Apr 2015)
14. Glazer, A.: Motivating devoted workers. International Journal of Industrial Organization 22(3), 427–440 (2004)
15. Goodman, J.K., Cryder, C.E., Cheema, A.: Data collection in a flat world: The strengths and weaknesses of mechanical turk samples. Journal of Behavioral Decision Making 26(3), 213–224 (2013)
16. Gualtieri, C.T., Johnson, L.G.: Reliability and validity of a computerized neurocognitive test battery, cns vital signs. Archives of Clinical Neuropsychology 21(7), 623–643 (2006)
17. Hartswood, M., Jirotka, M., Chenu-Abente, R., Hume, A., Giunchiglia, F., Martucci, L.A., Fischer-Hübner, S.: Privacy for peer profiling in collective adaptive systems. In: Privacy and Identity Management. IFIP Advances in Information and Communication Technology, vol. 457, pp. 237–252. Springer (2014)
18. Hauber, M., Bachmann, A., Budde, M., Beigl, M.: jActivity: Supporting mobile web developers with HTML5/JavaScript based human activity recognition. In: Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia. pp. 45:1–45:2. MUM '13, ACM, New York, NY, USA (2013)
19. Heckman, J.J., Smith, J.A., Taber, C.: What do bureaucrats do? The effects of performance standards and bureaucratic preferences on acceptance into the JTPA program. Advances in the Study of Entrepreneurship Innovation and Economic Growth 7, 191–217 (1996)
20. Hirth, M., Hoßfeld, T., Tran-Gia, P.: Anatomy of a crowdsourcing platform — Using the example of Microworkers.com. In: Workshop on Future Internet and Next Generation Networks (FINGNet). Seoul, Korea (Jun 2011)

21. Hirth, M., Scheuring, S., Hoßfeld, T., Schwartz, C., Tran-Gia, P.: Predicting result quality in crowdsourcing using application layer monitoring. In: Communications and Electronics (ICCE), 2014 Fifth International Conference on. IEEE (2014)
22. Hossfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K., Tran-Gia, P.: Best Practices for QoE Crowdtesting: QoE Assessment with Crowdsourcing. Transactions on Multimedia 16(2) (Feb 2014)
23. Irani, L.C., Silberman, M.S.: Turkopticon: Interrupting worker invisibility in Amazon Mechanical Turk. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 611–620. CHI '13, ACM, New York, NY, USA (2013)
24. Jacques, J.T., Kristensson, P.O.: Crowdsourcing a HIT: Measuring workers' pre-task interactions on microtask markets. In: First AAAI Conference on Human Computation and Crowdsourcing (Nov 2013)
25. Kazai, G., Zitouni, I.: Quality management in crowdsourcing using gold judges behavior. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. pp. 267–276. ACM (2016)
26. Kieffer, S., Dwyer, T., Marriott, K., Wybrow, M.: Hola: Human-like orthogonal network layout. IEEE Transactions on Visualization and Computer Graphics 22(1), 349–358 (Jan 2016)
27. Kittur, A., Nickerson, J.V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., Horton, J.: The future of crowd work. In: Proc. of the 2013 Conf. on Computer supported cooperative work. pp. 1301–1318. CSCW '13, ACM (2013)
28. Lebreton, P., Hupont, I., Mäki, T., Skodras, E., Hirth, M.: Eye tracker in the wild: Studying the delta between what is said and measured in a crowdsourcing experiment. In: Proceedings of the Fourth International Workshop on Crowdsourcing for Multimedia. pp. 3–8. ACM (2015)
29. Lebreton, P., Mäki, T., Skodras, E., Hupont, I., Hirth, M.: Bridging the gap between eye tracking and crowdsourcing. vol. 9394, pp. 93940W–93940W–14 (2015)
30. Little, G.: Exploring iterative and parallel human computation processes. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, Extended Abstracts Volume, Atlanta, Georgia, USA, April 10-15, 2010. pp. 4309–4314 (2010)
31. Mao, A., Kamar, E., Chen, Y., Horvitz, E., Schwamb, M.E., Lintott, C.J., Smith, A.M.: Volunteering versus work for pay: Incentives and tradeoffs in crowdsourcing. In: Hartman, B., Horvitz, E. (eds.) HCOMP. AAAI (2013)
32. Martin, D.B., Hanrahan, B.V., O'Neill, J., Gupta, N.: Being a Turker. In: Computer Supported Cooperative Work, CSCW '14, Baltimore, MD, USA, February 15-19, 2014. pp. 224–235 (2014)
33. Mason, W., Watts, D.J.: Financial incentives and the "performance of crowds". In: Proceedings of the ACM SIGKDD Workshop on Human Computation. pp. 77–85. HCOMP '09, ACM, New York, NY, USA (2009)
34. McDuff, D., el Kaliouby, R., Picard, R.W.: Crowdsourcing facial responses to online videos: Extended abstract. In: 2015 International Conference on Affective Computing and Intelligent Interaction (ACII). pp. 512–518 (Sep 2015)
35. McDuff, D., el Kaliouby, R., Picard, R.: Crowdsourced data collection of facial responses. In: Proceedings of the 13th International Conference on Multimodal Interfaces. pp. 11–18. ICMI '11, ACM, New York, NY, USA (2011)
36. Mok, R.K., Li, W., Chang, R.K.: Detecting low-quality crowdtesting workers. In: 2015 IEEE 23rd International Symposium on Quality of Service (IWQoS). pp. 201–206. IEEE (2015)
37. Narayanan, A., Shmatikov, V.: Robust de-anonymization of large sparse datasets. In: Security and Privacy, 2008. SP 2008. IEEE Symposium on. pp. 111–125. IEEE (2008)

38. Navalpakkam, V., Churchill, E.: Mouse tracking: Measuring and predicting users' experience of web-based content. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 2963–2972. CHI '12, ACM, New York, NY, USA (2012)
39. Peer, E., Samat, S., Brandimarte, L., Acquisti, A.: Beyond the turk: An empirical comparison of alternative platforms for online behavioral research. Available at SSRN 2594183 (Apr 2015)
40. Raddick, M.J., Bracey, G., Gay, P.L., Lintott, C.J., Murray, P., Schawinski, K., Szalay, A.S., Vandenberg, J.: Galaxy Zoo: Exploring the motivations of citizen science volunteers. Astronomy Education Review 9(1) (Dec 2010)
41. Reed, J., Raddick, M.J., Lardner, A., Carney, K.: An exploratory factor analysis of motivations for participating in Zooniverse, a collection of virtual citizen science projects. In: 2013 46th Hawaii International Conference on System Sciences (HICSS). pp. 610–619 (Jan 2013)
42. Richardson, D.W., Gribble, S.D.: Maverick: Providing web applications with safe and flexible access to local devices. In: In Proceedings of the 2011 USENIX Conference on Web Application Development (2011)
43. Salehi, N., Irani, L.C., Bernstein, M.S., Alkhatib, A., Ogbe, E., Milland, K., 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. pp. 1621–1630. CHI '15, ACM, New York, NY, USA (2015)
44. Shen, X.: Mobile crowdsourcing [Editor's note]. IEEE Network 29(3), 2–3 (May 2015)
45. Thuan, N.H., Antunes, P., Johnstone, D.: Factors influencing the decision to crowdsource: A systematic literature review. Information Systems Frontiers 18(1), 47–68 (Feb 2016)
46. Väätäjä, H.K., Ahvenainen, M.J., Jaakola, M.S., Olsson, T.D.: Exploring augmented reality for user-generated hyperlocal news content. In: CHI '13 Extended Abstracts on Human Factors in Computing Systems. pp. 967–972. CHI EA '13, ACM, New York, NY, USA (2013)
47. Vaillant, G.: Triumphs of Experience. Harvard University Press (2012)
48. Vakharia, D., Lease, M.: Beyond AMT: An analysis of crowd work platforms. In: iConference 2015 Proceedings. iSchools (Mar 2015)
49. Wu, H.Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., Freeman, W.: Eulerian video magnification for revealing subtle changes in the world. ACM Trans. Graph. 31(4), 65:1–65:8 (Jul 2012)
50. Xu, P., Ehinger, K.A., Zhang, Y., Finkelstein, A., Kulkarni, S.R., Xiao, J.: TurkerGaze: Crowdsourcing saliency with webcam based eye tracking. arXiv:1504.06755 [cs] (Apr 2015)