

INFLUENCING AND MEASURING BEHAVIOUR IN CROWDSOURCED ACTIVITIES

Sandy J.J. Gould¹, Anna L. Cox², Duncan P. Brumby²

¹University of Birmingham, ²University College London

Abstract

Crowdsourcing psychometric data is common in areas of Human-Computer Interaction (HCI) such as information visualization, text entry and interface design. In some of the social sciences, crowdsourcing data is now considered routine, and even standard. In this chapter, we explore the collection of data in this manner, beginning by describing the variety of approaches can be used to crowdsource data. Then, we evaluate past literature that has compared the results of these approaches to more traditional data-collection techniques. From this literature, we synthesize a set of design and implementation guidelines for crowdsourcing studies. Finally, we describe how particular analytic techniques can be recruited to aid the analysis of large-scale crowdsourced data. The goal of this chapter is to clearly enumerate the difficulties of crowdsourcing psychometric data and to explore how, with careful planning and execution, these limitations can be overcome.

Introduction

Crowdsourcing is a way for many activities to be completed remotely by distributed groups of people who often do not know each other and who, in most cases, will never meet. Completing activities in this way can allow for geographically diverse activities to take place (e.g., Biggs et al., 2015; Sullivan et al., 2009), or enable collaboration on tasks where expertise is sparsely distributed (e.g., Haklay, 2010).

Researchers are turning to crowds to answer research questions a variety of contexts, from finding out which birds are living where (Sullivan et al., 2009), to developing law reforms (Aitamurto et al., 2017). In some cases, crowds are used because they provide a way to quickly obtain data without having to consume the time and space that traditional laboratory studies often demand. For some kinds of research problems however, the only way that researchers can access the phenomena they want to study is to turn to crowds. This is because, working by themselves, researchers would be unable to obtain the breadth and depth of data they need to adequately address their research question.

In this chapter, we explore the methodological benefits and challenges of crowdsourcing data for HCI studies in this way. With a particular focus on the collection of psychometric data, we begin by outlining the different ways that researchers can collect data in the crowd, and assess the quality of data these methods produce. Building on our own experiences and those of other researchers, we explain how researchers can get the best out of crowds and how they should process the kinds of data that crowds produce. Our goal is to give readers who are new to crowdsourcing the tools that they need to avoid some of the most common pitfalls encountered when running a crowdsourced study. For experienced users of crowdsourcing platforms, we synthesize some of the latest research in the area and offer this chapter as an up-to-date desk reference on working with these platforms.

Types of crowdsourcing

The term ‘crowdsourcing’ is a very broad one that captures a wide range of activities. Some of this activity is related to the collection of research data. Much of it is not. Organisciak and Twidale (2015) developed a typology of crowdsourcing, in which they identify five main areas of crowdsourcing; *free and open source software*, *commons-based peer production*, *citizen science*, *the wisdom of crowds* and *human computation*. The focus of this chapter is on *crowdsourcing for research*, with citizen science and human computation the topics that we draw on most (rather than, for example, creating an online encyclopaedia or developing a new computer operating system). These topics are the most relevant to the collection of *psychometric data for 3rd Wave HCI*. We consider the two modes of crowdsourcing

production that are most relevant to these focuses: volunteer-based citizen science and paid crowdworking.

Volunteer-based citizen science

The practice of citizen science has existed for a long time. One of the earliest ‘citizen science’ projects (although it wouldn’t have been called ‘citizen science’ at the time), was the *Christmas Bird Count*. This bird-counting project started in 1900 (Silvertown, 2009). Historical citizen science projects, such as the Christmas Bird Project, involved writing down information and positing it, by mail, to a central location where researchers would aggregate the results by hand. This was a slow and laborious; citizen science has been revolutionised by the internet. While internet-based projects are properly known as *online* citizen science, the overwhelming majority of citizen science projects are now internet-based. ‘Citizen science’ now implies ‘online citizen science’.

Online citizen science projects usually involve researchers creating project websites, which might host sets of tasks for *contributors* to complete (e.g., labelling images), or data collection tools for them to submit observations (e.g., submitting an interesting plant that has been found). People visit these projects and contribute their time to complete tasks or make observations, and do so without any financial reward. Collaborators are often active participants in research, sometimes discovering new phenomena and appearing as authors on academic publications.

Paid crowdworking

‘Crowdworking’ is a term used to describe people who complete crowd-based tasks in exchange for money – hence *crowdworking*. Work on these platforms is usually broken down into small components (*microtasks*) that can be completed quickly and independently by a worker. (Although see Cheng et al., 2015 for a comparison of smaller and larger tasks.) Work is normally managed through crowdworking platforms, Amazon Mechanical Turk (or MTurk) being the most visible example. *Requesters* post the work to these platforms (which take a cut for their market-making) and *workers* see a list of available tasks, along with a requester-specified estimate completion time, and the amount of pay. Once a task has been completed by a worker, the requester assesses the quality of the work and approves payment. Generally, crowdworkers are participants only in data collection and generation – they are not normally involved in the latter stages of the research process (e.g., publications).

Comparative studies of crowdsourcing

Attitudes toward the use of crowdsourcing in scientific studies have changed significantly over the last few years. Even in research areas like Cognitive Science, which is less obviously connected to developments in crowdsourcing than HCI, crowdsourcing data is not only accepted, but normal (Stewart et al., 2017). Here we explore the evidential basis for substituting laboratory-based investigations with crowdsourced studies (i.e., whether data gathered from these different groups of participants has any effect of the conclusions that are drawn from the studies). We consider different types of research paradigms, explaining why some are likely to need more care and attention if they are to successfully transition to crowdsourcing platforms.

The challenges of making use of crowdsourcing in research vary significantly from discipline to discipline. Questionnaire-based studies have been run online for around twenty years (e.g., Wright, 2005), and have naturally moved onto crowdsourcing platforms (Behrend et al., 2011). Questionnaire-based studies have traditionally been run by physically posting questionnaires to people in remote locations, so moving a questionnaire study online makes little difference; the same challenges exist in terms of recruiting and giving instructions to remote participants regardless of whether the questionnaire is delivered through a postal form or an online form (Hoonakker and Carayon, 2009). Transitioning from postal methods to internet-based methods has therefore been easy: non-compliance, disappointing response rates and mindless responses are something that researchers using questionnaire-based techniques are used to dealing with in their traditional methods of data collection.

For experimental paradigms that have normally been based in laboratories, using crowds to obtain data marks a significant change. The core principle of an experiment is control. As many factors as possible should be fixed, leaving only experimental variables free to vary. The purpose of a laboratory is to offer as much control over confounding factors (e.g., computer hardware, lighting) as possible. When our experiments leave the lab and move into people's living rooms, or, say, onto public transport, we sacrifice control. When our experiments move from standard lab computers with standard lab screens and onto people's computers, phones and tablets, we sacrifice control.

Looser experimental control normally weakens the internal validity of an experiment. But precisely how confounding factors –uncontrolled influences on the dependent variables in our study– affect a study will depend on the kinds of measures that are being used. Some dependent measures are more easily confounded than others. It is important to consider how a given experimental paradigm might be affected by moving it onto an online crowdsourcing platform.

When considering the impact on data quality of running an experiment on a crowdsourcing platform, it is important to think about the types of data that are being collected. If a dependent variable is *categorical*, for instance if the measure is one choice from many, or if a solution produced is correct or not, a loss of experimental control may not be too problematic. For instance, Germine et al. (2012) compared lab-sourced and crowdsourced data from reasoning and decision making experiments where responses could easily be categorised as correct or incorrect. They found that crowdsourced data did not differ systematically from lab-sourced data.

In contrast, when an outcome variable is in the form of *ratio* data, deploying the experimental paradigm online is more challenging. Timing data is a good example of the kind of ratio data we might commonly collect. Psychometric studies frequently use reaction times or task completion times as performance measures. Time-sensitive experiments –where time is a dependent variable– might be particularly susceptible to the loss of control that comes with moving experiments into the crowd. Comparisons of crowdsourced and lab-sourced data of this kind would be the most likely to give us a clearer picture of the limits of crowdsourcing.

Komarov et al. (2013) compared crowd- and lab-sourced data from several menu-search experiments. Menu searching is a common practical HCI problem that has received significant research attention (e.g., Bailly et al., 2014; Brumby et al., 2014; Brumby and Howes, 2008; Brumby and Zhuang, 2015). Experiments typically involve making small adjustments to the presentation of menus and then measuring search times to see if those adjustments have affected performance. Reaction times are typically in the low thousands of milliseconds, so even small disturbances of a couple of hundred milliseconds have the potential to distort results. The results of Komarov et al.'s comparisons show that data collected from Amazon Mechanical Turk workers is statistically equivalent to those collected from laboratory participants. In other words, even in their time-sensitive experiments, crowd data were as good as laboratory data.

We have also explored the challenges of using time-based measures in our own work. In Gould, Cox, Brumby and Wiseman (2015b) we compared timing data from a multitasking experiment that was run in a laboratory with timing data from the same task run on a crowdsourcing platform. We found that although there were absolute differences in the data that we gathered, these differences between the online and lab-based timing data were not statistically significant.

Touchscreen input is another substantial research area in HCI that makes use of fine-grained psychometric data (e.g., Dunlop and Levine, 2012; Oulasvirta et al., 2013). It is a particularly challenging style of research to conduct on crowdsourcing platforms because of the variety of devices people use, many of which have differently sized screens and different pixel densities. Findlater et al. (2017) compared laboratory- and crowd-collected data from an investigation of touchscreen interactions. When they analysed the results from the two sources independently, they came to two different conclusions about what their data showed. When they looked at the data in more detail, they

found that this was because crowdworkers were generally faster and less accurate in their pointing. They suggest that researchers should be cautious about using the crowd for these kinds of touchscreen pointing experiments.

As well as considering the type of data that is being collected, researchers should also need think about what the subject of that data will be. In some research paradigms, participants provide data about the world (e.g., Kim et al., 2011). In such studies, participants act as sensors reporting things about the world. For example, in the eBird project (Sullivan et al., 2009), contributors record the species of birds they have seen in a particular area. Other projects have used crowdsourcing platforms to report driving offences (Aubry et al., 2014). In others, particularly psychological studies, participants are themselves data, whether through their reporting of their attitudes or by objective recording of their behaviour. Jennett et al. (2014) described this form of participant-as-data study as ‘citizen psychscience’.

The distinction between data that is from sources endogenous or exogenous to the participants is important when thinking about the potential challenges of crowdsourced research in 3rd Wave HCI contexts. We know from research in personal informatics that people frequently give-up on self-tracking tools, which involve the logging of person-generated (‘endogenous’) data (Epstein et al., 2016). This is often because people do not form habits for this kind of logging (Stawarz et al., 2015). Researchers looking to crowdsource this kind of personal tracking information must consider the challenges of sustaining participation in these contexts.

Experimental control is, of course, a trade-off. Often in search of greater control and internal validity, we end up sacrificing ecological validity by making contrived tasks developed by the experimenter for the sole purpose of the experiment. The evidence we have considered so far suggests that crowdsourcing is a good option for many, but not all, experimental paradigms. Researchers must consider the threats to internal validity that running an experiment in the crowd can bring and the options available to them for preserving control and bolstering validity. In the next section, we explore how empirical studies can best be designed for crowdsourcing platforms.

Designing empirical studies for crowdsourcing platforms

Much of the comparative literature on lab-based and crowd-based studies yields insights into how experiments can be moved from more traditional experimental media and onto paid and volunteer crowdsourcing platforms. In this section, we consider these findings, integrating them with other insights from the literature to explain how best to design empirical studies for deployment on crowdsourcing platforms. We consider which platforms researchers should choose, the challenges of recruitment, and designing for the human aspects of participation.

Choosing a platform

The capacity of crowd platforms to produce results at scale quickly is predicated on having a large pool of potential workers to call on. If researchers want to get results quickly and in volume, they should use one of the large crowdworking platforms. Amazon Mechanical Turk claims to have over 500,000 active workers (see Peer et al., 2017). For citizen science projects, researchers might try to partner with one of the larger platforms (e.g., Zooniverse) to find potential collaborators.

Platform selection is not as simple as choosing the biggest platform. Peer et al. (2017) investigated two other paid crowdworking platforms, *Crowdfunder* and *Prolific Academic*. They focused on whether different platforms would produce the same results. Results of their comparative study found that Amazon Mechanical Turk workers are particularly well drilled in completing experiments – and are excellent at coping with the mechanisms researchers use to try and ensure participants’ compliance with experimental procedures. Workers on Crowdfunder were the least likely to spot instructional manipulation checks the researchers inserted (see later section on Attention Checks), and also produced the lowest quality responses, on average. For research where participant naivety is critical for the success of an experiment, switching to platforms where workers are less experienced (and therefore ‘streetwise’ about experiments) might be a good idea.

As well as having to think about which crowd platform to target (e.g., Galaxy Zoo, Mechanical Turk), researchers must also consider which hardware platforms they are trying to target. Is a researcher's expectation that people will participate sitting at a desk, using a traditional keyboard and mouse? Lots of crowdworking happens on more traditional computing devices, but people now have all sorts of other devices, like smartphones and tablets, many of which they use for participation, perhaps while out and about.

In many ways, mobile computing has typified changes in technology corresponding to research efforts in the 3rd Wave of HCI (Bødker, 2015). Around the world, many people's experience of the internet is solely through mobile technology (Gitau et al., 2010). Mobile technology has changed the way that we work (Cecchinato et al., 2015a) and play (Colley et al., 2017). People's mobile technologies, such as smartphones and smartwatches, are a constant source of interruption and distraction (Cecchinato et al., 2017, 2015; Pielot et al., 2014; Rigby et al., 2017). Designing for this context, especially in a crowdsourcing context where attention might be especially limited, means keeping tasks short and interruptible.

Some researchers have taken advantage of the fact we spend a lot of our day unlocking our phones (Harbach et al., 2016), realising that rather than entering meaningless codes or swipes, we could unlock our phones by doing something useful (Truong et al., 2014). Researchers have realised that lockscreens might be a good location for the siting of the kinds of very small microtasks that crowdworkers often work on. Vaish et al. (2014) implemented 'Twitch', a platform for crowdsourcing microtasks on smartphone lockscreens. Working on tasks that took less than two seconds each to complete, the 82 participants in their field study completed over eleven thousand tasks.

Not all psychometric experiments can be completed in the space of a couple of seconds, but phones can still provide a platform for more time-consuming studies. Brown et al. (2014) investigated the potential of phones for crowdsourcing data in cognitive science studies. They developed a game-based environment for running experiments and deployed it as a smartphone app (see Gray, 2017 for a broader discussion). The app was used to investigate a number of psychological phenomena, from working memory to attentional blink. Across large samples, they showed that classic findings from cognitive psychology could be replicated through a game-based app that crowdsourced participants.

Smartphones are not just used as mechanisms by which participants submit data about bird counts, or platforms for interactive crowdsourcing tasks. The variety of sensors on modern devices and the ease with which applications giving access to these sensors can be installed on mobile devices means that people can participate in complex sensing studies by doing nothing more than installing an application (Mehrotra et al., 2016). This lowers the barriers to participation in studies that would otherwise be onerous for participants.

Researchers should think carefully before they develop a bespoke smartphone application for a crowdsourcing project. Unlike web applications, mobile apps, particularly ones with complex sensing functionality are not very portable – it may be necessary to develop multiple applications for different platforms (Brown et al., 2014). One way to make development less resource intensive is to build plugins for existing app-based platforms for deploying studies (e.g., Chatzimilioudis et al., 2012). Platforms like the *Aware Framework* (Ferreira et al., 2015) make it relatively simple for researchers to collect and aggregate significant amounts of sensor data. These data can be augmented with other platform-provided data, like, for instance, people's responses to user experience reports. Installing a sensing applications on devices can also be deleterious to a device's performance, making it essential that such sensing applications are as efficient as possible (Lane et al., 2013); lest a slower device and reduced battery life change people's behaviour and influence researchers' conclusions.

In general, researchers should avoid developing purpose-built applications for studies. If simply having to sign-up to contribute puts potential participants off (Jay et al., 2016), then having to download and install a new app is likely to have a similar effect on levels of participation. There needs to be a

compelling reason to crowdsource data through dedicated native applications (e.g., extremely time-sensitive studies, the need for sensor data) for them to be worth the effort of implementing over a simple website.

Recruiting participants

For those designing studies for deployment on crowd platforms, it is important to remember that crowd platforms are not just collections of independent individuals. Crowdsourcing platforms usually have communities behind them (see, e.g., Irani and Silberman, 2013; Moore et al., 2011; Reeves et al., 2017). This is often advantageous to requesters and researchers recruiting from crowdsourcing platforms. Existing communities can be a source of well-motivated contributors (Preece, 2016). The new communities that form around projects also develop long-term understandings of projects that help contributors perform, for instance, data validation tasks (Wiggins and He, 2016).

In paid crowdworking, ‘communities’ are usually constituted by the financial incentives offered by crowdworking platforms. For volunteer-based citizen science, however, developing a new community is the biggest challenge for most projects. No community often means no contributors, and no contributors means no citizen science project. In the absence of financial incentives, understanding the motivations of volunteer citizen scientists becomes more important. The challenge of building communities is common across a number of 3rd Wave HCI contexts, from the organization of physical communities to political action. Starting new communities is very difficult, but even trying to involve existing communities (e.g., Crivellaro et al., 2016; Vlachokyriakos et al., 2016) is a hard problem to solve. It can help to understand how individual characteristics mediate behaviours. For instance, people’s willingness to involve themselves in civic activities (e.g., volunteering, attending neighbourhood meetings) is influenced by personality traits (Kim et al., 2013); for more introverted people social media plays a greater role in their participation in such activities. The internet is a powerful tool for engaging with people who might be missed by more traditional ways of contributing.

The better the motivations of potential contributors to citizen science projects are understood, the greater the chance of a community coalescing around a project. Jennett and colleagues (2016) looked across the motivations of citizen scientists documented in the literature. One of the most common motivators was a prior interest in the research topic, which suggests successful citizen science projects will tap into existing interests; developing people’s interest in a topic while trying to get them to contribute might be difficult. Additionally, Jennett et al. found that people’s motivations change over time, so researchers should not assume that contributors who are well motivated will stay well motivated. It is also important to understand the kinds of people that might be attracted to citizen science projects. Not everyone will be interested in contributing to citizen science projects. Contributors to projects may not have science as part of their everyday life, but they tend to have a high level interest in science, and strongly value the contribution of science to broader society (Land-Zandstra et al., 2015).

Having a strong community is mostly beneficial, but there are some important drawbacks of having a community with shared experiences completing similar (or the same) tasks. Information about tasks is back-channelled through communities through tools like message boards. This presents challenges for recruiters. The extent to which the community component of crowd platforms can be a help or hindrance depends on the kinds of activities a requester wants to implement, and the extent to which they engage with and support their communities.

We often assume that participants in our studies are naïve. That is to say that we do not expect our participants to know what the research is about, and how exactly we have operationalized our measures or what tricks and shortcuts there are to get tasks completed as quickly as possible. If participants know these things before they participate, they are more likely to exhibit *demand characteristics*, which is to say that they will express the attitudes that a researcher desires, or they will approach a task in the way that a researcher hopes. This is not desirable, because it increases the chances of the results of a study leading to Type I (false positive) errors.

Different experimental paradigms will vary in the extent to which they are affected by participants already knowing the purpose of a study and mechanisms by which it operates. Econometric studies, for instance (e.g., Rand, 2012), are particularly susceptible to this issue. Often, economists are interested in how people make decisions about how to allocate resources between themselves and a collaborator. The capacity for such studies to tell us something about people's reasoning is contingent on the experiments being 'closed systems' – participants should only have the information that is made available to them in the experimental task. Contributors to crowdsourced projects are frequently *not* naïve, however (Chandler et al., 2014). This is because information about the purpose of a study may have been posted in public space, such as a message board. Private out-of-system collusion in collaborative tasks can also skew results. The non-naivety of participants is problematic, because it reduces experiments' chances of detecting effects (Chandler et al., 2015). Researchers must account for this kind of back-channelling behaviour in their studies, particularly if they are running experimental paradigms that crowdworkers frequently encounter on their platforms (Stewart et al., 2017).

Not all non-naivety is a problem. In some instances, 'naivety' represents the first step on a learning curve. The goal might be to complete a given task as efficiently as possible, so the quicker that a contributor can get 'up to speed', the sooner they will be producing useful results. Classification tasks, such as the ones hosted on Galaxy Zoo (Lintott et al., 2008), are a good example of such an activity. Volunteers look at (often fuzzy) images of galaxies and stars, and aid researchers by classifying these images. It takes time for new volunteers to understand which categories exist and how the noisy data they are given should be slotted into these categories. Being naïve in this context does not aid the research – having significant knowledge of the target problem does not inhibit the quality of responses from contributors. In this context, non-naivety is something that's helpful, because it's a skilled task that takes time to learn. Supporting new volunteers with tools is therefore essential so that they are properly taught how to complete the task to a good standard. That said, it is also important to recognize that one of the advantages of citizen science is that it brings in people with different and new perspectives. Sometimes these new perspectives help to correct errors in perceived wisdom. One example of this is Galaxy Zoo's 'Green Peas' (Cardamone et al., 2009). Volunteers on the project noticed a particular kind of star appearing from time-to-time. These kinds of stars had already been classified by experts. But the citizen scientists persevered, discussing their findings on the Galaxy Zoo message boards. Ultimately their resistance to perceived wisdom was what allowed them to realise the existing classification was incomplete, allowing for a new discovery to occur. So, it is important that the learning process, while enabling people to make contributions, does not destroy these valuable additional perspectives.

Designing for real participants

The nature of crowdsourcing platforms can often lull researchers into thinking that their tasks are completed by anonymous units of production rather than real individuals. Experiments are sent into the ether; results magically appear within moments. But it is important to remember the people who provide the workforce that crowd platforms rely on. Remembering that they are people rather than units of production will improve the quality of results that are returned.

Crowdwork is based on the idea of task decomposition. Tasks can be broken down into small units that can be independently completed. While the content of two distinct units of work might be independent, they are not necessarily completed by independent workers: workers string together multiple tasks, one after another (or often at the same time). Newell and Ruths (2016) investigated how one crowdworking task can interfere with behaviour on another. They found significant 'intertask effects', where the cognitive style demanded by one task caused participants to behave differently in a completely separate task because of the priming effects of the initial task. For instance, an image labelling task on a given topic might prime behaviour on another image labelling task. Or the framing of a particular labelling task (i.e., the scenario) affects the subsequent completion of the task itself, even though the framing is not functionally related to the task.

Another important practical thing to consider is fatigue. Crowdworkers typically spend four hours a day on crowd tasks, and up to five-and-a-half hours if crowdsourcing is their primary source of income (Lasecki et al., 2015). This means that many independent tasks are worked on back-to-back. Fatigue is one of the potential side effects of this kind of working, so researchers have looked at whether inserting breaks into streams of microtasks can help to improve performance. Rzeszotarski et al. (2013) looked at matching microtasks with *microbreaks* – very short breaks inserted between batches of tasks. They found that inserting microbreaks improved worker retention, and also improved overall worker engagement and satisfaction. This is important for requesters, as retaining high quality workers who are familiar with a task improves the quality of results returned. Other more extensive work has also looked at the effect of breaks on performance (e.g., Dai et al., 2015), and come to similar conclusions – giving workers a chance to recuperate improves their performance. For long tasks made up of batches of tasks, researchers should consider giving workers a break.

Once a participant starts working on a task that has been set, the other tasks that they need to work on do not disappear. Other activities that they are working on still need to be worked on. For example, people receive messages from friends or colleagues that need responding to, people have caring responsibilities, and people may have their ‘main’ job to attend to at the same time. This means that people in the crowd frequently switch to other tasks while they work (Gould et al., 2013, 2015a, 2016a). There are a class of psychometric studies, particularly those studies that test working memory, where such distractions might significantly impact results. While people can be induced to keep focused on their task (Gould et al., 2016a), it is not possible to eliminate such distractions altogether. Instead, researchers might find it better to record switching events, use this measure when processing their data. (See the discussion of data handling later in the chapter.)

Collecting data using crowdsourcing platforms

Building and continuing engagement

Encouraging participants to engage in paid crowdwork is usually straightforward: money motivates participants, additional money motivates further participation. In the context of volunteer-based crowdsourcing, the question of how to engage potential participants and convert them into regular, high-quality contributors is one of the most important research questions in the area, and probably the most important practical question for a researcher looking to set-up a crowd-based citizen science study.

When people are volunteering their time, any barrier between finding out about a project and contributing to it is likely to mean potential contributors are not converted to actual contributors. One example of a barrier that might turn potential contributors off from a project is having to sign up for an account before they get a chance to see if the activities provided by a project are ones that they’d like to get involved with. Jay et al. (2016) found that removing the requirement that people create an account before participating increased participation by 62%. Participants who subsequently want to claim their contributions can create an account at their convenience. The idea of reducing barriers to entry fits with Eveleigh et al.’s idea of ‘designing for dabblers’ (2014). In the course of their research, they found that not everyone is a potential significant contributor. Many people are satisfied with participating a little and then moving on to something else. If all contributions are useful to a given project (an important proviso), then encouraging these kinds of contributions, without putting too much pressure on people to become regular contributors, can help to bolster participation levels.

Ethical data collection

All experiments involving human participants need to be conducted in an ethical manner, and in compliance with local regulations. Crowdsourcing data brings a number of special ethical considerations, however. Much of the ethical debate in the context of paid crowdworking has centred around payment and working practices. Sometimes these issues range from the obvious, such as people not being paid for the work they have done, to more fundamental issues with the way that online labour systems are organized.

The working conditions of crowdworkers have come under significant scrutiny. Silberman (2010) reported in detail on the issues that crowdworkers face. A lot has been written about the low rates of pay that crowdworks often suffer, with the average worker earning a median of USD\$2 per hour (Hara et al., 2017). This is not the only issue. From the cost of requester errors being borne by workers (e.g., tasks being set-up incorrectly, or having bugs in them) to simple non-payment (i.e., wage theft), or being drawn into platform-hosted scams, there are many potential barriers to crowdworkers getting paid for the work that they have done.

There has been some movement towards improving conditions by some requesters. Kittur et al. (2013) ask how crowdwork might be turned into something “*which we would want our children to participate*” in, and enumerate a number of suggestions for making this a reality. Some are focused on improving technical aspects of crowdworking, so that errors are less likely to occur. Other suggestions include ways of developing better understandings of crowdworkers so that requesters can create tasks with which they are more likely to succeed.

One of the challenges for crowdworkers is that they are individuals in unequal power relationships with the requesters they are working for. There is normally little recourse if the crowdworker if they are cheated by requesters, but researchers and workers have made some efforts to correct this imbalance. Irani and Sliberman (2013) set-up the ‘Turkopticon’ platform in an attempt to give workers more power in the system, and augment existing channels for worker-worker collaborations, such as message boards. Turkopticon allows workers to report their activity on tasks in Amazon Mechanical Turk. Workers can rate requesters on a number of attributes, from whether the pay was fair, to whether communication from a requester was accurate. Although it has no official role and does not arbitrate any disputes, it is a valuable resource: workers can check on requesters before they agree to work for them. The motivation for this was to reduce their chances of being cheated. A more ambitious idea to allow crowdworkers to take collective action was ‘Dynamo’ (Salehi et al., 2015), which helped workers engage in collective action over time without losing momentum when disagreements in how to proceed were encountered. An alternative strategy to helping workers regain some control from platform owners and work requesters is to simply put workers in control of the crowd platforms (Fox et al., 2017).

Many ethical review boards are still not well adapted to crowd-based forms of research (Graber and Graber, 2012). Nevertheless, researchers should still do their best to make sure that workers are treated fairly and compensated for the work that they do, making sure that they respond to the legitimate concerns of workers in good time.

Citizen science projects are not necessarily in a better ethical position by virtue of not involving the exchange of money. There are a host of ethical concerns that come along with citizen science projects. One of these is privacy, and the exposure of contributing citizen scientists’ personal information. Bowser et al. (2017) point to location privacy as a particularly important issue: participants in projects like *eBird* tag their sightings of birds with longitude and latitude co-ordinates. Across a number of postings, it is possible to work out both spatial and temporal components of people’s movements. Bowser et al. report one contributor to a citizen science project writing that “*someone will inadvertently put a comment to say, “temperature was 79 degrees, and by the way this is my last report for the next week because I’m going out of town”*” (Bowser et al., 2017, p. 2131). Of course, between this information and the contributor’s contribution history on the project, some potentially undesirable people could work out where their unoccupied house was. Triangulating multiple sources of data – location-tagged contributions, message board posts, social media posts means people end up revealing a lot more about themselves than they might have intend. It is important when building citizen science studies to consider the trade-offs between community building and the potential for personal information leakage.

Resnik et al. (2015) reflect on the ethical concerns of other authors (e.g., Riesch and Potter, 2014) in questioning whether citizen science has the potential to be exploitative, given the lack of obvious remuneration in many citizen science projects. We discuss this issue further in the next section on remuneration, but Resnik et al.’s point is that inequitable exchanges between researchers and research

participants (as contributors are often characterized) are unethical – participants need to receive something back from their participation. Resnik et al. suggest a variety of strategies to ensure collaborators receive something back depending on the project and the nature of the contribution that contributors have made. This might range from authorship to certificates to education on the topic they have contributed to.

There is also the question of whether, as co-investigators on projects, citizen scientists themselves have ethical obligations, such as the disclosure of conflicts of interest. We are not aware of any citizen science projects that formally place ethical demands on contributors as well as the researchers running programmes. For some research projects where citizen scientists are involved in data collection and the topic involves sensitive subjects (e.g., the location of certain endangered species), having a degree of ethical training available to citizen scientists might be advisable.

Payment and non-monetary remuneration

Paid crowdsourcing platforms have historically been described as ‘cheap’ and ‘cost effective’. Compared to the time and effort of running psychometric experiments in laboratories this is true. We have already considered some of the ethical issues that surround payment (and non-payment) in crowdsourcing settings. Here we will focus instead on the more practical aspects of how participants respond to incentives, and how the choice of incentives that are used affects participants’ behaviours. Although issues of remuneration are most pertinent in the context of paid crowdworking, we also consider how non-monetary remuneration affects behaviour in volunteer crowdsourcing contexts.

It would be easy to assume that the relationship between pay and quality in crowdworking contexts is trivial: the more workers are paid, the better the quality of the work that they return. In fact, while the amount paid for a task is an important factor in the contributions a researcher receives, the relationship between pay and remuneration in crowdworking settings is nuanced. Prior research has found, for example, that while workers might contribute less when they are paid less, they do not work less if they are given a particularly time-consuming task for the same amount of money (Horton and Chilton, 2010). The same authors found that people tend to target certain amounts of money when working, and are happy to stop at that point, irrespective of whether that is the most ‘rational’ strategy at a given moment.

A number of studies have attempted to understand the relationship between pay and quality (e.g., Hsieh et al., 2010; Mason and Watts, 2010; Shaw et al., 2011). In an early study, Mason and Watts (2010) found that increasing financial incentives for work generally increased the quantity of work people would produce, but did not necessarily increase the quality of the work produced. Other studies have replicated this finding (e.g., Hsieh et al., 2010). The size of the financial incentive is, therefore, not always the limiting factor on performance. Combining incentives with other approaches, like getting participants to think about how their peers might respond (Shaw et al., 2011) can improve performance. Rather than have poor-performing workers continue to contribute low quality data that subsequently has to be removed from analyses, sometimes it is more effective to pay poorly performing workers not to contribute any further, leaving tasks for more competent workers (Harris, 2015).

It is important to remember that although financial incentives explain why many people get involved in crowdwork, it is not the only reason, nor is the only factor that people take into account when they are choosing which work to do. Jiang et al. (2015) found that monetary compensation might be the main motivator, but that workers are also motivated by learning something new and feeling fulfilled by contributing to a bigger task.

Factors besides money can also be the primary motivation for participants. In previous work we have explicitly compared monetary and non-monetary reimbursement for participants (Wiseman et al., 2017). One of the things we wanted to study was whether participants could be motivated by being given data about themselves as an outcome of a study. This approach has proved to be successful on platforms like *Lab in the Wild* (Reinecke et al., 2013; Reinecke and Gajos, 2015), where people are given feedback on, for instance, their memory capacity. This has produced some very large snowballed

samples, where participants recruit other participants. In Wiseman et al. (2017), we compared this kind of data-as-reward approach with other approaches, like simply requesting help for the good of science or in return for payment. We found that that participants in both non-monetary conditions met, and in many cases exceeded, the performance of participants in our control paid condition. Other work has found similar patterns, with participants rewarded with information about themselves performing better than those rewarded with money (Ye et al., 2017). These findings suggest that in the absence of monetary compensation, researchers need to be more creative with the kinds of rewards that they use to keep people involved and engaged.

Ensuring quality data

Experiments can be designed in a way that maximizes the quality of data that they produce. There are two major reasons why otherwise well-designed experimental studies might produce poor quality results. First, participants do not follow the instructions for the experiment, either because they do not read or because they do not understand the instructions. The second is that, participants, whether they understand the instructions or not, do not comply with them.

There are two stages to having participants assimilate instructions for an experiment. The first is to convince participants that they should actually read the instructions. The second is to deliver instructions in such a way that people are able to understand them and act in accordance with them. Getting participants to actually read the instructions that accompany an experiment is also important for ethical reasons. To give properly informed consent, participants need to have read and understood what a study is for and how the data they produce will be used. Getting participants to engage with instructions yields both better data and is a requirement for research ethics.

Satisficing, where people optimize their behaviour to gain maximum return on their effort, is a major challenge in all experiments, but it is particularly problematic in paid online experiments. Paid work on crowdsourcing platforms are usually offered in a piecework fashion (see Alkhatib et al., 2017 for more background), which means workers are paid for each task, rather than how long they spend on it. This creates an incentive to get through a task as quickly as possible, and so data quality can suffer.

Attention checks

One common way to try and deal with inattentiveness and satisficing –in experiments of all types– is to make use of *attention checks*. These comprise a variety of methods, but the goal is the same – to detect when participants have not understood the instructions for a study or when their attention has wandered while they are completing it. Kapelner and Chandler (2010) conducted one of the first investigations of satisficing –where people look to get as much return for as little effort as possible– in crowdworking settings. They looked specifically at satisficing by crowdworkers as they completed questionnaires. The temptation for some workers might be to click through the questionnaire as quickly as possible, so that they can finish and claim their payment. This rushing –however well intentioned– has the potential to generate poor quality data. Removing poor quality responses from a sample requires additional researcher time and effort, so such responses are worse than useless. Kapelner and Chandler developed a variety of approaches to reduce satisficing behaviour. One intervention simply involved introducing a small delay between the presentation of a question and the means to answer. Building on this design, they introduced a ‘Kapcha’, which, on top of the delay in presenting response widgets, also faded the text of the questionnaire into view at a rate of 250 words per minute – this prevented participants from skim-reading. This easily implemented intervention improved the quality of responses by 10%.

Attention checks can be used *correctively* (to encourage re-engagement) or they can be used *surreptitiously* to aid the discovery and handling of poor quality results. Abbey and Meloy (2017) produced a detailed overview of a number of different types of attention checks, and reviewed their efficacy. They found that, on average, attention checks of all kinds improved the quality of results that studies obtained. In the following paragraphs review the variety of attention checks Abbey and Meloy review.

Attention checks have been used in traditional psychological research for some time. Abbey and Meloy (Abbey and Meloy, 2017) point to comments made by Rensis Likert –the inventor of the Likert scale– about the necessity of using reversed questions in questionnaires. Reversed questions in questionnaires are duplicate questions but with the meaning of the question inverted. For instance, if a question on a scale were “I always feel down about my prospects”, the reversed question would be “I never feel down about my prospects”. If participants report strongly agreeing with both of these statements, it is likely that they are not giving their full attention to the questions being asked of them.

One type of attention check that Abbey and Meloy (2017) describe in their review are *instructional manipulation checks*. Instructional manipulation checks are elements inserted into an experiment that can help researchers to be more certain that participants have read and understood the experimental instructions. For example, Oppenheimer, Meyvis and Davidenko (2009) constructed survey questions in a way that meant participants could not skim over them. Participants would be faced with an introductory paragraph, followed by a question and several clickable responses. Participants who had read the instructions as requested would have found in the text instructions to ignore the question and click on the title of the page instead.

Corrigan-Gibbs et al. (2015) investigated what Abbey and Meloy call ‘Honesty Checks’. They created a survey that was about the domain name of a particular website they told participants was being developed. Participants in one condition were asked to follow an ‘honor code’ when completing the survey, which meant agreeing not to ask other people about the survey or to visit other websites for information about the task. Participants were required to retype the honor code. Instead of the honor code, participants in another condition saw a warning text that listed three repercussions of visiting other websites, including non-payment. In this condition participants had to type a one-sentence summary of the warning.

Unbeknownst to the participants, the authors were closely monitoring activity on the domain that participants had been told not to visit; they were able to match-up participants in Amazon Mechanical Turk with visitors to the website. In this way, cheating could be detected. Corrigan-Gibbs et al. found that the harsh warning was significantly more effective than the honor code in discouraging cheating, although the honor code was more effective than the control condition, where no message relating to cheating was shown. The result is not surprising. Workers make a living on Amazon Mechanical Turk, so anything that negatively affects the metrics that represent them on the platform, for instance the number of tasks they have completed, or their rejection rates, materially affects their chances of getting work (see, e.g., Silberman et al., 2010).

Interestingly, Corrigan-Gibbs et al. initially required participants to record themselves reading the honor code aloud. This extra step substantially increased the drop-out rate from the study, suggesting that attention checks need to be carefully calibrated so that they not radically deter participation.

We explored the efficacy of attention checks in one of our own studies. In Gould, Cox and Brumby (2016a), we have also looked at how task-switching behaviour during the completion of a task on Amazon Mechanical Turk influenced overall task performance. We wondered whether participants could be encouraged to stick to the task at hand, rather than switching to other activities while they completed our task. Our approach was not aligned with any of Abbey and Meloy’s (2017) different types of attention checks, because it was *dynamically sensitive to participants’ behaviour*. We detected when participants switched away to other tasks. When participants returned to our task, they were either met with a notice asking them to stay focused on our task or a question asking them what they had been doing. Participants in the control condition received no indication that their switching behaviour had been detected. We found that participants who were asked to focus on the task at hand were significantly less likely to allow themselves to be interrupted. Given that participants who switched less performed best, our attention check may have improved the performance of participants in that condition.

Attention checks are not suitable for all kinds of study. Researchers investigating certain phenomenon should consider whether the design of their study is suitable for the deployment of instructional manipulation checks. Hauser and Schwarz (2015) ran an experiment on Amazon Mechanical Turk that investigated how the inclusion of instructional manipulation checks affect behaviour later on, *during the experimental task*. They found that participants who had encountered an instructional manipulation check exhibited increased levels of systematic thinking on decision making tasks. Hauser and Schwarz's explanation was that the instructional manipulation check caused participants to become warier of the experiment as a whole; participants were primed by the instruction manipulation checks to be on the lookout for 'tricky' elements in the main experiment. Consequently, the presence of the check affected participants performance in the rest of the experiment.

Most of the evidence we have comes from studies of paid workers. Satisficing among volunteer citizen scientists is likely to take a different form, as there is no financial reward for volunteers to be had from getting through a task as quickly as possible. Participants might, for example, be satisficing to improve their contribution statistics. There are, to our knowledge, no empirical investigations of satisficing in the context of citizen science research, however, so this is an area that future work could investigate.

Participant preselection

Even if participants read and understand instructions, and then try their best to follow instructions, their performance might still be poor. Some participants will just find certain tasks difficult. In some experimental paradigms, this kind of poor performance is specifically of interest, for example in investigations of individual differences in behaviour (e.g., Meys and Sanderson, 2013). In other experimental paradigms, reaching a certain threshold of competence is necessary before participants start producing useful data, for example in studies investigating typing performance on novel keyboards (e.g., Oulasvirta et al., 2013). If prospective participants cannot meet the threshold required to produce useful data, it might be best to filter them out of the process early on.

Developing preselection processes can help to ensure that participants with certain behavioural characteristics are channelled into the study, while those potential participants who do not can be filtered out. On crowdsourcing platforms like Amazon Mechanical Turk, ideal candidates can be granted qualifications that give them access to the full experiment. It is also possible to predict how a participant will perform on a task based using generic measures of behaviours that are required across a broad range of tasks. Rzeszotarski and Kittur (2011) showed that a collection of standard measures (e.g., typing speed, inter-task time duration), predicts performance in a variety of distinct tasks. Rather than assessing participants on task completion, dispositional factors might also be useful in understanding participants' performance. Bored participants, for example, are less likely to be pay attention to details in tasks (Jun et al., 2017). Thus, tasks that require great attention to detail may benefit from deselecting people who say they are participating because they are bored; they may be better directed to other tasks that are less sensitive to inattentiveness.

An alternative to using experimenter-administered tests is to use participant assessment. In the context of paid crowdworking participants have a strong incentive not to accept tasks that they will not be able to complete to a sufficiently good standard. If it is of poor quality, workers will not be paid, and the cost is effectively doubled once the opportunity cost of working on one task over another is factored-in. However, as Gadiraju et al. (2017) point out, even when well-motivated self-assessment are not always accurate, because, as psychological evidence attests less competent participants are less likely to be able accurately assess their abilities (e.g., Dunning, 2011; Dunning et al., 2004). In an empirical study on Amazon Mechanical Turk, Gadiraju et al. indeed found that the least competent workers overestimated their ability. To compensate, Gadiraju et al. combined self-assessments with more objective measures, such as results from standard logical reasoning tasks, to weed-out the participants who were poor estimators of their performance. Gadiraju et al. found that, with the help of their tools for supporting self-assessment, workers who struggled with the task deselected themselves. This allowed more skilled workers to complete more tasks. This improved the overall accuracy of the sample by more than 15%.

Analysing data from crowdsourcing platforms

Some experimental paradigms do not lend themselves very well to instruction manipulation or attention checks. Participants might rush through a study, or researchers might overlook the potential for participants to misunderstand the instructions for a task. Regardless, the outcome is the same – the results of studies contain a greater or lesser degree of poor quality data.

Once a dataset has been assembled, identifying and correcting for poor quality results should be the first stage in an analytical pipeline. Detecting and dealing with outliers is a standard feature of psychological studies, but in laboratory-sourced data, researchers might have some ideas about which participants may not have performed well before they start their analyses. There are no hints from observation in crowdsourced studies, so researchers must rely entirely on the data they collected. Xu et al. (2013) looked at how outliers could be detected and removed from crowdsourced experiments, and whether removing outliers improved the quality of results that were obtained. They developed the *LASSO* technique, which automatically filtered outliers based on rigorous statistical measures built on the concept of *robust regression*. This improved the overall quality of the datasets they were creating.

Researchers should consider the characteristics of their own research methods when applying such techniques. In subjective rating tasks, for example, shaping responses using outlier detection can have the effect of homogenising responses that do not need homogenisation (Riegler et al., 2016). When it comes to subjective responses in tasks without ‘gold standard’ benchmarks, removing outliers might actually decrease the overall quality of data by removing alternative perspectives. It is therefore important that outlier detection is paired with the other techniques we have described to determine whether outlying responses are the result of non-compliance with instructions or simply a different perspective on the task at hand.

As well as looking directly at the dependent measures that a study has collected to identify poor quality data, collecting additional telemetry can also help researchers identify participants whose performance is hard to explain. In our own work, we have shown how crowdworkers’ propensity to switch to other tasks predicts time-based measures of their performance on a task (Gould et al., 2016b, 2016a, 2015a). Browser events, like whether a window has lost focus can give a researcher an indication of how distracted a participant has been. Other measures, like clicks and taps or keyboard interactions can also help build up a picture of how a participant has interacted with a task. The measures can either be used to help clean-up poor quality entries in a data set or, as we have discussed previously, be aggregated at a larger scale to predict performance and pre-select participants before any time or money is wasted (Rzeszotarski and Kittur, 2011).

Automated statistical processes can help analyse data quickly and efficiently. They also require a well-developed procedure and a clear idea of what thresholding criteria for removal will be used. Sometimes, particularly when using a new experimental paradigm, it is necessary to ‘eyeball’ the data collected to look for patterns or potential outliers across the measures that experimental telemetry can yield. Visualizing data can help with this process. Rzeszotarski and Kittur (2012) developed *CrowdScape* with this in mind. *CrowdScape* comprises a suite of different visualization tools that allow behaviour data to be quickly consumed. For instance, in a task that required users to type relatively long responses, a plot of the aggregated keypress data for each participants would allow a researcher to very quickly see which participants had contributed significant amounts, and which participants had not typed very much. Through this process of visual exploration, researchers can either identify outliers, or, if they are being more systematic, allow them to develop formal selection criteria for outliers.

Ultimately the benefits of crowdsourcing do not come at zero cost. If a tiny minority of participants does their best to circumvent our best efforts to ensure quality data, we do not have unlimited power to detect them. An extra degree of noise is inevitable and is an intrinsic part of the method.

Summary

At the start of this chapter, we aimed to “*synthesize a set of design and implementation guidelines for crowdsourcing studies*”. The most important things to take-away from this chapter are:

- Generally, data from crowd platforms is at least as good as data collected in the lab.
- The success in translating a psychometric study from a lab to a crowd will depend on the type of data collected and what the collected data is about. Not all studies are equally amenable to being run online.
- Attention checks are a useful tool for keeping participants focused on the task at hand, but participants in crowdsourced studies are often alert to the possibility that attention checks may be present. Researchers should ensure that attention checks are not too easy for participants to circumvent.
- Naïve participants are sometimes important, sometimes a hindrance. Think about whether a study will be affected by prior-knowledge. If not, focus on retaining participants who have achieved a good level of competence.
- Paying participants well is important, but it is not the only thing to consider. Good quality instructions can help people produce good quality work.
- Non-monetary reimbursement, perhaps in the form of personal data, can be an even better motivator than money.
- Crowdworkers are people trying to make a living, not anonymous units of internet production. They get tired. One task can blur into the next. Microtasks are not truly independent.
- It is easier to stop bad data getting into a sample than it is to try and remove it afterwards.

Conclusion

Thanks to the significant efforts of researchers across a variety of domains, crowdsourcing methods have rapidly matured. Many of the practical challenges we have identified in crowdsourcing are similar to ones that other researchers have encountered working on other 3rd Wave HCI problems, and there is potential for successful strategies to transcend narrow research topics. Diverse platforms now have large numbers of potential participants, and techniques for getting the best out of those participants are increasingly well developed and understood. Data collection and analysis techniques have been developed to help researchers improve the quality of the data they collect. On the whole, crowdsourcing platforms work well for most forms of data collection. In some experimental paradigms, crowdsourcing might be superior to more traditional data collection techniques, not just in quantity and quality, but in the kinds of questions that such platforms allow researchers to explore.

In this chapter we have covered a wide range of crowdsourcing literature to find out what works best when conducting a study on a crowdsourcing platform. We have given particular focus to crowdsourcing in the context of the types of experiments (like psychometric studies) that might be particularly affected by the transition from the tightly controlled environment of the laboratory to the more chaotic and unpredictable world of crowdsourcing. Our goal has been to provide a synthesis that means other researchers nip common issues in the bud, rather than have to deal with them after significant amounts of empirical data collection have already taken place.

References

Abbey, J.D., Meloy, M.G., 2017. Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *J. Oper. Manag.* <https://doi.org/10.1016/j.jom.2017.06.001>

- Aitamurto, T., Landemore, H., Galli, J.S., 2017. Unmasking the crowd: participants' motivation factors, expectations, and profile in a crowdsourced law reform. *Inf. Commun. Soc.* 20, 1239–1260. <https://doi.org/10.1080/1369118X.2016.1228993>
- Alkhatib, A., Bernstein, M.S., Levi, M., 2017. Examining Crowd Work and Gig Work Through The Historical Lens of Piecework, in: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17*. ACM, New York, NY, USA, pp. 4599–4616. <https://doi.org/10.1145/3025453.3025974>
- Aubry, E., Silverston, T., Lahmadi, A., Festor, O., 2014. CrowdOut: A mobile crowdsourcing service for road safety in digital cities, in: *2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS)*. Presented at the 2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS), pp. 86–91. <https://doi.org/10.1109/PerComW.2014.6815170>
- Bailly, G., Oulasvirta, A., Brumby, D.P., Howes, A., 2014. Model of Visual Search and Selection Time in Linear Menus, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14*. ACM, New York, NY, USA, pp. 3865–3874. <https://doi.org/10.1145/2556288.2557093>
- Behrend, T.S., Sharek, D.J., Meade, A.W., Wiebe, E.N., 2011. The viability of crowdsourcing for survey research. *Behav. Res. Methods* 43, 800–813. <https://doi.org/10.3758/s13428-011-0081-0>
- Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Dejean, T., Griffiths, R.A., Foster, J., Wilkinson, J.W., Arnell, A., Brotherton, P., Williams, P., Dunn, F., 2015. Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). *Biol. Conserv.*, Special Issue: Environmental DNA: A powerful new tool for biological conservation 183, 19–28. <https://doi.org/10.1016/j.biocon.2014.11.029>
- Bødker, S., 2015. Third-wave HCI, 10 Years Later—participation and Sharing. *interactions* 22, 24–31. <https://doi.org/10.1145/2804405>
- Bowser, A., Shilton, K., Preece, J., Warrick, E., 2017. Accounting for Privacy in Citizen Science: Ethical Research in a Context of Openness, in: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17*. ACM, New York, NY, USA, pp. 2124–2136. <https://doi.org/10.1145/2998181.2998305>
- Brown, H.R., Zeidman, P., Smittenaar, P., Adams, R.A., McNab, F., Rutledge, R.B., Dolan, R.J., 2014. Crowdsourcing for Cognitive Science – The Utility of Smartphones. *PLOS ONE* 9, e100662. <https://doi.org/10.1371/journal.pone.0100662>
- Brumby, D.P., Cox, A.L., Chung, J., Fernandes, B., 2014. How Does Knowing What You Are Looking for Change Visual Search Behavior?, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14*. ACM, New York, NY, USA, pp. 3895–3898. <https://doi.org/10.1145/2556288.2557064>
- Brumby, D.P., Howes, A., 2008. Strategies for Guiding Interactive Search: An Empirical Investigation Into the Consequences of Label Relevance for Assessment and Selection. *Human-Computer Interact.* 23, 1–46. <https://doi.org/10.1080/07370020701851078>
- Brumby, D.P., Zhuang, S., 2015. Visual Grouping in Menu Interfaces, in: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*. ACM, New York, NY, USA, pp. 4203–4206. <https://doi.org/10.1145/2702123.2702177>
- Cardamone, C., Schawinski, K., Sarzi, M., Bamford, S.P., Bennert, N., Urry, C.M., Lintott, C., Keel, W.C., Parejko, J., Nichol, R.C., Thomas, D., Andreescu, D., Murray, P., Raddick, M.J., Slosar, A., Szalay, A., VandenBerg, J., 2009. Galaxy Zoo Green Peas: discovery of a class of compact extremely star-forming galaxies. *Mon. Not. R. Astron. Soc.* 399, 1191–1205. <https://doi.org/10.1111/j.1365-2966.2009.15383.x>
- Cecchinato, M.E., Cox, A.L., Bird, J., 2017. Always On(Line)?: User Experience of Smartwatches and Their Role Within Multi-Device Ecologies, in: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17*. ACM, New York, NY, USA, pp. 3557–3568. <https://doi.org/10.1145/3025453.3025538>
- Cecchinato, M.E., Cox, A.L., Bird, J., 2015. Working 9-5?: Professional Differences in Email and Boundary Management Practices, in: *Proceedings of the 33rd Annual ACM Conference on*

- Human Factors in Computing Systems, CHI '15. ACM, New York, NY, USA, pp. 3989–3998. <https://doi.org/10.1145/2702123.2702537>
- Chandler, J., Mueller, P., Paolacci, G., 2014. Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behav. Res. Methods* 46, 112–130. <https://doi.org/10.3758/s13428-013-0365-7>
- Chandler, J., Paolacci, G., Peer, E., Mueller, P., Ratliff, K.A., 2015. Using Nonnaive Participants Can Reduce Effect Sizes. *Psychol. Sci.* 0956797615585115. <https://doi.org/10.1177/0956797615585115>
- Chatzimilioudis, G., Konstantinidis, A., Laoudias, C., Zeinalipour-Yazti, D., 2012. Crowdsourcing with Smartphones. *IEEE Internet Comput.* 16, 36–44. <https://doi.org/10.1109/MIC.2012.70>
- Cheng, J., Teevan, J., Iqbal, S.T., Bernstein, M.S., 2015. Break It Down: A Comparison of Macro- and Microtasks, in: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*. ACM, New York, NY, USA, pp. 4061–4064. <https://doi.org/10.1145/2702123.2702146>
- Colley, A., Thebault-Spieker, J., Lin, A.Y., Degraen, D., Fischman, B., Häkkinen, J., Kuehl, K., Nisi, V., Nunes, N.J., Wenig, N., Wenig, D., Hecht, B., Schöning, J., 2017. The Geography of Pokémon GO: Beneficial and Problematic Effects on Places and Movement, in: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17*. ACM, New York, NY, USA, pp. 1179–1192. <https://doi.org/10.1145/3025453.3025495>
- Corrigan-Gibbs, H., Gupta, N., Northcutt, C., Cutrell, E., Thies, W., 2015. Deterring Cheating in Online Environments. *ACM Trans Comput-Hum Interact* 22, 28:1–28:23. <https://doi.org/10.1145/2810239>
- Crivellaro, C., Taylor, A., Vlachokyriakos, V., Comber, R., Nissen, B., Wright, P., 2016. Re-Making Places: HCI, “Community Building” and Change, in: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*. ACM, New York, NY, USA, pp. 2958–2969. <https://doi.org/10.1145/2858036.2858332>
- Dai, P., Rzeszotarski, J.M., Paritosh, P., Chi, E.H., 2015. And Now for Something Completely Different: Improving Crowdsourcing Workflows with Micro-Diversions, in: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15*. ACM, New York, NY, USA, pp. 628–638. <https://doi.org/10.1145/2675133.2675260>
- Dunlop, M., Levine, J., 2012. Multidimensional Pareto Optimization of Touchscreen Keyboards for Speed, Familiarity and Improved Spell Checking, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*. ACM, New York, NY, USA, pp. 2669–2678. <https://doi.org/10.1145/2207676.2208659>
- Dunning, D., 2011. *The Dunning-Kruger Effect: On Being Ignorant of One's Own Ignorance*, in: *Advances in Experimental Social Psychology*. Academic Press.
- Dunning, D., Heath, C., Suls, J.M., 2004. Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychol. Sci. Public Interest J. Am. Psychol. Soc.* 5, 69–106. <https://doi.org/10.1111/j.1529-1006.2004.00018.x>
- Epstein, D.A., Caraway, M., Johnston, C., Ping, A., Fogarty, J., Munson, S.A., 2016. Beyond Abandonment to Next Steps: Understanding and Designing for Life After Personal Informatics Tool Use, in: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*. ACM, New York, NY, USA, pp. 1109–1113. <https://doi.org/10.1145/2858036.2858045>
- Eveleigh, A., Jennett, C., Blandford, A., Brohan, P., Cox, A.L., 2014. Designing for Dabblers and Deterring Drop-outs in Citizen Science, in: *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14*. ACM, New York, NY, USA, pp. 2985–2994. <https://doi.org/10.1145/2556288.2557262>
- Ferreira, D., Kostakos, V., Dey, A.K., 2015. AWARE: Mobile Context Instrumentation Framework. *Front. ICT* 2. <https://doi.org/10.3389/fict.2015.00006>
- Findlater, L., Zhang, J., Froehlich, J.E., Moffatt, K., 2017. Differences in Crowdsourced vs. Lab-based Mobile and Desktop Input Performance Data, in: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17*. ACM, New York, NY, USA, pp. 6813–6824. <https://doi.org/10.1145/3025453.3025820>

- Fox, S., Dimond, J., Irani, L., Hirsch, T., Muller, M., Bardzell, S., 2017. Social Justice and Design: Power and Oppression in Collaborative Systems, in: Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17 Companion. ACM, New York, NY, USA, pp. 117–122. <https://doi.org/10.1145/3022198.3022201>
- Gadiraju, U., Fetahu, B., Kawase, R., Siehndel, P., Dietze, S., 2017. Using Worker Self-Assessments for Competence-Based Pre-Selection in Crowdsourcing Microtasks. *ACM Trans Comput-Hum Interact* 24, 30:1–30:26. <https://doi.org/10.1145/3119930>
- Germine, L., Nakayama, K., Duchaine, B.C., Chabris, C.F., Chatterjee, G., Wilmer, J.B., 2012. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychon. Bull. Rev.* 19, 847–857. <https://doi.org/10.3758/s13423-012-0296-9>
- Gitau, S., Marsden, G., Donner, J., 2010. After Access: Challenges Facing Mobile-only Internet Users in the Developing World, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10. ACM, New York, NY, USA, pp. 2603–2606. <https://doi.org/10.1145/1753326.1753720>
- Gould, S.J.J., Cox, A.L., Brumby, D.P., 2016a. Diminished Control in Crowdsourcing: An Investigation of Crowdworker Multitasking Behavior. *ACM Trans Comput-Hum Interact* 23, 19:1–19:29. <https://doi.org/10.1145/2928269>
- Gould, S.J.J., Cox, A.L., Brumby, D.P., 2015a. Task lockouts induce crowdworkers to switch to other activities. *Conf. Hum. Factors Comput. Syst. - Proc.* 18, 1785–1790.
- Gould, S.J.J., Cox, A.L., Brumby, D.P., 2013. Frequency and Duration of Self-initiated Task-switching in an Online Investigation of Interrupted Performance, in: Human Computation and Crowdsourcing: Works in Progress and Demonstration Abstracts AAAI Technical Report CR-13-01. Presented at the First AAAI Conference on Human Computation and Crowdsourcing, AAAI, pp. 22–23.
- Gould, S.J.J., Cox, A.L., Brumby, D.P., Wickersham, A., 2016b. Now Check Your Input: Brief Task Lockouts Encourage Checking, Longer Lockouts Encourage Task Switching, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16. ACM, New York, NY, USA, pp. 3311–3323. <https://doi.org/10.1145/2858036.2858067>
- Gould, S.J.J., Cox, A.L., Brumby, D.P., Wiseman, S., 2015b. Home is Where the Lab is: A Comparison of Online and Lab Data From a Time-sensitive Study of Interruption. *Hum. Comput.* 2, 45–67. <https://doi.org/10.15346/hc.v2i1.4>
- Graber, M.A., Graber, A., 2012. Internet-based crowdsourcing and research ethics: the case for IRB review. *J. Med. Ethics* medethics-2012-100798. <https://doi.org/10.1136/medethics-2012-100798>
- Gray, W.D., 2017. Game-XP: Action Games as Experimental Paradigms for Cognitive Science. *Top. Cogn. Sci.* 9, 289–307. <https://doi.org/10.1111/tops.12260>
- Haklay, M., 2010. How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environ. Plan. B Plan. Des.* 37, 682–703. <https://doi.org/10.1068/b35097>
- Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., Bigham, J., 2017. A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk. *ArXiv171205796 Cs*.
- Harbach, M., De Luca, A., Egelman, S., 2016. The Anatomy of Smartphone Unlocking: A Field Study of Android Lock Screens, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16. ACM, New York, NY, USA, pp. 4806–4817. <https://doi.org/10.1145/2858036.2858267>
- Harris, C.G., 2015. The Effects of Pay-to-Quit Incentives on Crowdworker Task Quality, in: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15. ACM, New York, NY, USA, pp. 1801–1812. <https://doi.org/10.1145/2675133.2675185>
- Hauser, D.J., Schwarz, N., 2015. It's a Trap! Instructional Manipulation Checks Prompt Systematic Thinking on “Tricky” Tasks. *SAGE Open* 5, 2158244015584617. <https://doi.org/10.1177/2158244015584617>
- Hoonakker, P., Carayon, P., 2009. Questionnaire Survey Nonresponse: A Comparison of Postal Mail and Internet Surveys. *Int. J. Hum.-Comput. Interact.* 25, 348. <https://doi.org/10.1080/10447310902864951>

- Horton, J.J., Chilton, L.B., 2010. The Labor Economics of Paid Crowdsourcing, in: Proceedings of the 11th ACM Conference on Electronic Commerce, EC '10. ACM, New York, NY, USA, pp. 209–218. <https://doi.org/10.1145/1807342.1807376>
- Hsieh, G., Kraut, R.E., Hudson, S.E., 2010. Why Pay?: Exploring How Financial Incentives Are Used for Question & Answer, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10. ACM, New York, NY, USA, pp. 305–314. <https://doi.org/10.1145/1753326.1753373>
- Irani, L.C., Silberman, M.S., 2013. Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13. ACM, New York, NY, USA, pp. 611–620. <https://doi.org/10.1145/2470654.2470742>
- Jay, C., Dunne, R., Gelsthorpe, D., Vigo, M., 2016. To Sign Up, or Not to Sign Up?: Maximizing Citizen Science Contribution Rates Through Optional Registration, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16. ACM, New York, NY, USA, pp. 1827–1832. <https://doi.org/10.1145/2858036.2858319>
- Jennett, C., Furniss, D.J., Iacovides, I., Wiseman, S., Gould, S.J.J., Cox, A.L., 2014. Exploring Citizen Psych-Science and the Motivations of Errordriary Volunteers. *Hum. Comput. 1*. <https://doi.org/10.15346/hc.v1i2.10>
- Jennett, C., Kloetzer, L., Schneider, D., Iacovides, I., Cox, A., Gold, M., Fuchs, B., Eveleigh, A., Methieu, K., Ajani, Z., Talsi, Y., 2016. Motivations, learning and creativity in online citizen science. *J. Sci. Commun. 15*.
- Jiang, L., Wagner, C., Nardi, B., 2015. Not Just in it for the Money: A Qualitative Investigation of Workers' Perceived Benefits of Micro-task Crowdsourcing, in: 2015 48th Hawaii International Conference on System Sciences (HICSS). Presented at the 2015 48th Hawaii International Conference on System Sciences (HICSS), pp. 773–782. <https://doi.org/10.1109/HICSS.2015.98>
- Jun, E., Hsieh, G., Reinecke, K., 2017. Types of Motivation Affect Study Selection, Attention, and Dropouts in Online Experiments. *Proc ACM Hum-Comput Interact 1*, 56:1–56:15. <https://doi.org/10.1145/3134691>
- Kapelner, A., Chandler, D., 2010. Preventing Satisficing in online surveys, in: CrowdConf.
- Kim, S., Robson, C., Zimmerman, T., Pierce, J., Haber, E.M., 2011. Creek Watch: Pairing Usefulness and Usability for Successful Citizen Science, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11. ACM, New York, NY, USA, pp. 2125–2134. <https://doi.org/10.1145/1978942.1979251>
- Kim, Y., Hsu, S.-H., de Zúñiga, H.G., 2013. Influence of Social Media Use on Discussion Network Heterogeneity and Civic Engagement: The Moderating Role of Personality Traits. *J. Commun. 63*, 498–516. <https://doi.org/10.1111/jcom.12034>
- Kittur, A., Nickerson, J.V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., Horton, J., 2013. The future of crowd work, in: Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW '13. ACM, New York, NY, USA, pp. 1301–1318. <https://doi.org/10.1145/2441776.2441923>
- Komarov, S., Reinecke, K., Gajos, K.Z., 2013. Crowdsourcing Performance Evaluations of User Interfaces, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13. ACM, New York, NY, USA, pp. 207–216. <https://doi.org/10.1145/2470654.2470684>
- Land-Zandstra, A.M., Devilee, J.L.A., Snik, F., Buurmeijer, F., Broek, J.M. van den, 2015. Citizen science on a smartphone: Participants' motivations and learning. *Public Underst. Sci. 0963662515602406*. <https://doi.org/10.1177/0963662515602406>
- Lane, N.D., Chon, Y., Zhou, L., Zhang, Y., Li, F., Kim, D., Ding, G., Zhao, F., Cha, H., 2013. Piggyback CrowdSensing (PCS): Energy Efficient Crowdsourcing of Mobile Sensor Data by Exploiting Smartphone App Opportunities, in: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, SenSys '13. ACM, New York, NY, USA, p. 7:1–7:14. <https://doi.org/10.1145/2517351.2517372>
- Lasecki, W.S., Rzeszutarski, J.M., Marcus, A., Bigham, J.P., 2015. The Effects of Sequence and Delay on Crowd Work, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in

- Computing Systems, CHI '15. ACM, New York, NY, USA, pp. 1375–1378. <https://doi.org/10.1145/2702123.2702594>
- Lintott, C.J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M.J., Nichol, R.C., Szalay, A., Andreescu, D., Murray, P., Vandenberg, J., 2008. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Mon. Not. R. Astron. Soc.* 389, 1179–1189. <https://doi.org/10.1111/j.1365-2966.2008.13689.x>
- Mason, W., Watts, D.J., 2010. Financial incentives and the “performance of crowds.” *SIGKDD Explor Newsl* 11, 100–108. <https://doi.org/10.1145/1809400.1809422>
- Mehrotra, A., Pejovic, V., Vermeulen, J., Hendley, R., Musolesi, M., 2016. My Phone and Me: Understanding People’s Receptivity to Mobile Notifications, in: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*. ACM, New York, NY, USA, pp. 1021–1032. <https://doi.org/10.1145/2858036.2858566>
- Meys, H.L., Sanderson, P.M., 2013. The Effect of Individual Differences on How People Handle Interruptions. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 57, 868–872. <https://doi.org/10.1177/1541931213571188>
- Moore, J., Gay, P.L., Hogan, K., Lintott, C., Impey, C., Watson, C., 2011. Facebooking Citizen Science with the Zooniverse, in: *Bulletin of the American Astronomical Society*. Presented at the *Bulletin of the American Astronomical Society*, p. 15813.
- Newell, E., Ruths, D., 2016. How One Microtask Affects Another, in: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*. ACM, New York, NY, USA, pp. 3155–3166. <https://doi.org/10.1145/2858036.2858490>
- Oppenheimer, D.M., Meyvis, T., Davidenko, N., 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *J. Exp. Soc. Psychol.* 45, 867–872. <https://doi.org/10.1016/j.jesp.2009.03.009>
- Organisciak, P., Twidale, M.B., 2015. *Design Facets of Crowdsourcing*.
- Oulasvirta, A., Reichel, A., Li, W., Zhang, Y., Bachynskyi, M., Vertanen, K., Kristensson, P.O., 2013. Improving Two-thumb Text Entry on Touchscreen Devices, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13*. ACM, New York, NY, USA, pp. 2765–2774. <https://doi.org/10.1145/2470654.2481383>
- Peer, E., Brandimarte, L., Samat, S., Acquisti, A., 2017. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *J. Exp. Soc. Psychol.* 70, 153–163. <https://doi.org/10.1016/j.jesp.2017.01.006>
- Pielot, M., Church, K., de Oliveira, R., 2014. An In-situ Study of Mobile Phone Notifications, in: *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services, MobileHCI '14*. ACM, New York, NY, USA, pp. 233–242. <https://doi.org/10.1145/2628363.2628364>
- Preece, J., 2016. Citizen Science: New Research Challenges for Human–Computer Interaction. *Int. J. Human–Computer Interact.* 32, 585–612. <https://doi.org/10.1080/10447318.2016.1194153>
- Rand, D.G., 2012. The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *J. Theor. Biol., Evolution of Cooperation* 299, 172–179. <https://doi.org/10.1016/j.jtbi.2011.03.004>
- Reeves, N., Tinati, R., Zerr, S., Van Kleek, M.G., Simperl, E., 2017. From Crowd to Community: A Survey of Online Community Features in Citizen Science Projects, in: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17*. ACM, New York, NY, USA, pp. 2137–2152. <https://doi.org/10.1145/2998181.2998302>
- Reinecke, K., Gajos, K.Z., 2015. LabintheWild: Conducting Large-Scale Online Experiments With Uncompensated Samples, in: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15*. ACM, New York, NY, USA, pp. 1364–1378. <https://doi.org/10.1145/2675133.2675246>
- Reinecke, K., Nguyen, M.K., Bernstein, A., Näf, M., Gajos, K.Z., 2013. Doodle Around the World: Online Scheduling Behavior Reflects Cultural Differences in Time Perception and Group Decision-making, in: *Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW '13*. ACM, New York, NY, USA, pp. 45–54. <https://doi.org/10.1145/2441776.2441784>

- Resnik, D.B., Elliott, K.C., Miller, A.K., 2015. A framework for addressing ethical issues in citizen science. *Environ. Sci. Policy* 54, 475–481. <https://doi.org/10.1016/j.envsci.2015.05.008>
- Riegler, M., Gaddam, V.R., Larson, M., Eg, R., Halvorsen, P., Griwodz, C., 2016. Crowdsourcing as self-fulfilling prophecy: Influence of discarding workers in subjective assessment tasks, in: 2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI). Presented at the 2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI), pp. 1–6. <https://doi.org/10.1109/CBMI.2016.7500256>
- Riesch, H., Potter, C., 2014. Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Underst. Sci.* 23, 107–120. <https://doi.org/10.1177/0963662513497324>
- Rigby, J.M., Brumby, D.P., Gould, S.J.J., Cox, A.L., 2017. Media Multitasking at Home: A Video Observation Study of Concurrent TV and Mobile Device Usage, in: Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video, TVX '17. ACM, New York, NY, USA, pp. 3–10. <https://doi.org/10.1145/3077548.3077560>
- Rzeszotarski, J.M., Chi, E., Paritosh, P., Dai, P., 2013. Inserting Micro-Breaks into Crowdsourcing Workflows, in: Human Computation and Crowdsourcing: Works in Progress and Demonstration Abstracts AAAI Technical Report CR-13-01. Presented at the First AAAI Conference on Human Computation and Crowdsourcing, AAAI, pp. 62–63.
- Rzeszotarski, J.M., Kittur, A., 2012. CrowdScape: interactively visualizing user behavior and output, in: Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology, UIST '12. ACM, New York, NY, USA, pp. 55–62. <https://doi.org/10.1145/2380116.2380125>
- Rzeszotarski, J.M., Kittur, A., 2011. Instrumenting the crowd: using implicit behavioral measures to predict task performance, in: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST '11. ACM, New York, NY, USA, pp. 13–22. <https://doi.org/10.1145/2047196.2047199>
- Salehi, N., Irani, L.C., Bernstein, M.S., Alkhatib, A., Ogbe, E., Milland, K., Clickhappier, 2015. We Are Dynamo: Overcoming Stalling and Friction in Collective Action for Crowd Workers, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15. ACM, New York, NY, USA, pp. 1621–1630. <https://doi.org/10.1145/2702123.2702508>
- Shaw, A.D., Horton, J.J., Chen, D.L., 2011. Designing Incentives for Inexpert Human Raters, in: Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work, CSCW '11. ACM, New York, NY, USA, pp. 275–284. <https://doi.org/10.1145/1958824.1958865>
- Silberman, M.S., Irani, L., Ross, J., 2010. Ethics and Tactics of Professional Crowdsourcing. *XRDS* 17, 39–43. <https://doi.org/10.1145/1869086.1869100>
- Silvertown, J., 2009. A new dawn for citizen science. *Trends Ecol. Evol.* 24, 467–471. <https://doi.org/10.1016/j.tree.2009.03.017>
- Stawarz, K., Cox, A.L., Blandford, A., 2015. Beyond Self-Tracking and Reminders: Designing Smartphone Apps That Support Habit Formation, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15. ACM, New York, NY, USA, pp. 2653–2662. <https://doi.org/10.1145/2702123.2702230>
- Stewart, N., Chandler, J., Paolacci, G., 2017. Crowdsourcing Samples in Cognitive Science. *Trends Cogn. Sci.* 21, 736–748. <https://doi.org/10.1016/j.tics.2017.06.007>
- Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S., 2009. eBird: A citizen-based bird observation network in the biological sciences. *Biol. Conserv.* 142, 2282–2292. <https://doi.org/10.1016/j.biocon.2009.05.006>
- Truong, K.N., Shihpar, T., Wigdor, D.J., 2014. Slide to X: Unlocking the Potential of Smartphone Unlocking, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14. ACM, New York, NY, USA, pp. 3635–3644. <https://doi.org/10.1145/2556288.2557044>
- Vaish, R., Wyngarden, K., Chen, J., Cheung, B., Bernstein, M.S., 2014. Twitch Crowdsourcing: Crowd Contributions in Short Bursts of Time, in: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14. ACM, New York, NY, USA, pp. 3645–3654. <https://doi.org/10.1145/2556288.2556996>

- Vlachokyriakos, V., Crivellaro, C., Le Dantec, C.A., Gordon, E., Wright, P., Olivier, P., 2016. Digital Civics: Citizen Empowerment With and Through Technology, in: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '16. ACM, New York, NY, USA, pp. 1096–1099. <https://doi.org/10.1145/2851581.2886436>
- Wiggins, A., He, Y., 2016. Community-based Data Validation Practices in Citizen Science, in: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, CSCW '16. ACM, New York, NY, USA, pp. 1548–1559. <https://doi.org/10.1145/2818048.2820063>
- Wiseman, S., Cox, A.L., Gould, S.J.J., Brumby, D.P., 2017. Exploring the effects of non-monetary reimbursement for participants in HCI research. *Hum. Comput.* 4, 1–24. <https://doi.org/10.15346/hc.v4i1.1>
- Wright, K.B., 2005. Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services. *J. Comput.-Mediat. Commun.* 10, 00–00. <https://doi.org/10.1111/j.1083-6101.2005.tb00259.x>
- Xu, Q., Xiong, J., Huang, Q., Yao, Y., 2013. Robust Evaluation for Quality of Experience in Crowdsourcing, in: Proceedings of the 21st ACM International Conference on Multimedia, MM '13. ACM, New York, NY, USA, pp. 43–52. <https://doi.org/10.1145/2502081.2502083>
- Ye, T., Reinecke, K., Robert, L.P., Jr., 2017. Personalized Feedback Versus Money: The Effect on Reliability of Subjective Data in Online Experimental Platforms, in: Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17 Companion. ACM, New York, NY, USA, pp. 343–346. <https://doi.org/10.1145/3022198.3026339>