Exploring the effects of non-monetary reimbursement for participants in HCI research

SARAH WISEMAN, Goldsmiths, University Of London

ANNA L. COX, University College London

SANDY J.J. GOULD, University College London

DUNCAN P. BRUMBY, University College London

ABSTRACT

When running experiments within the field of Human Computer Interaction (HCI) it is common practice to ask participants to come to a specified lab location, and reimburse them monetarily for their time and travel costs. This, however, is not the only means by which to encourage participation in scientific study. Citizen science projects, which encourage the public to become involved in scientific research, have had great success in getting people to act as sensors to collect data or to volunteer their idling computer or brain power to classify large data sets across a broad range of fields including biology, cosmology and physical and environmental science. This is often done without the expectation of payment. Additionally, data collection need not be done on behalf of an external researcher; the Quantified Self (QS) movement allows people to reflect on data they have collected about themselves. This too, then, is a form of non-reimbursed data collection. Here we investigate whether citizen HCI scientists and those interested in personal data produce reliable results compared to participants in more traditional lab-based studies. Through six studies, we explore how participation rates and data quality are affected by recruiting participants without monetary reimbursement: either by providing participants with data about themselves as reward (a QS approach), or by simply requesting help with no extrinsic reward (as in citizen science projects). We show that people are indeed willing to take part in online HCI research in the absence of extrinsic monetary reward, and that the data generated by participants who take part for selfless reasons, rather than for monetary reward, can be as high quality as data gathered in the lab and in addition may be of higher quality than data generated by participants given monetary reimbursement online. This suggests that large HCI experiments could be run online in the future, without having to incur the equally large reimbursement costs alongside the possibility of running experiments in environments outside of the lab.

1. INTRODUCTION

Experiments in the field of Human Computer Interaction (HCI) can help researchers evaluate the way that people interact with devices. Controlled lab experiments can provide researchers with a wealth of data about the timing and accuracy of participants' interaction with a particular aspect of a technology. By running the same task many times, with small, systematic adjustments it is possible to establish potential causal links between particular features of technology and human behavior. However, the controlled nature of the lab often results in experiments that are long, and sometimes tedious for the participants to take part in and for the researcher to administer. In order to incentivize participation, monetary reimbursements are often offered. As experiments become larger and require more participants, this can mean the costs of conducting the research can get prohibitively large, meaning smaller sample sizes are sometimes used, resulting in lower data quality (Bertamini & Munafo, 2012). Finding alternative methods to incentivize participation in HCI experiments could therefore improve data quality.

The citizen science (CitSci) movement has shown that it is possible to encourage participants to engage in research without remunerating participants (Bonney, Ballard, et al., 2009). In citizen science projects, research groups ask members of the public to get involved with scientific research in a number of ways: from recording data from their environment (e.g. (The Royal Society for the Protection of Birds, 2015)), to transcribing old manuscripts (e.g. (Old Weather, 2015)), to classification of existing data (e.g. (Galaxy Zoo, 2015)). In the case of transcription and classification tasks, researchers are able to produce online software (sometimes framed as games, or "Games with a Purpose" (Von Ahn & Dabbish, 2008)) that can help volunteers to contribute to a program of research in some way. Members of the public are then invited to interact with the project. They are able to help move the research forward and together contribute a volume of man-hours that would be impossible for a traditional research team. This approach has led to scientific breakthroughs, for instance the citizen science project "Fold-It" (Fold It, 2015) has resulted in a significant discovery in the field of molecular biology (Khatib et al., 2011). As these examples demonstrate, there are benefits to be had from using citizen scientists in research – work can be completed in a fraction of the time and at minimal cost in comparison to running lab studies and in addition can raise awareness of research programs occurring in academic establishments that otherwise may not have high profile awareness outside of the academic community.

The Quantified Self movement (QS) has also shown how data collection need not be monetarily incentivized. QS involves the collection of data about oneself; a classic example of this is monitoring step count using a pedometer (Swan, 2013). With this data, the user is then able to reflect upon their own habits and lifestyle, which can then lead to behavior change (Bravata et al., 2007). The scope of data that can be collected is growing rapidly, with users able to track their sleep, water intake and even sneezes (Curtis, 2014). Unlike CitSci, the motivation for collection of data comes not from the "selfless" act of participating in an experiment belonging to someone else, but rather from the opportunity to learn more about oneself.

Both CitSci and QS encourage participants to become involved in data collection without monetary reimbursement (although for different reasons). It seems possible therefore that these approaches may be beneficial for HCI research. However, it is not immediately clear how they could be applied to this domain. Citizen science projects have generally used volunteers as interchangeable labor resources that can be applied to solve a larger problem. In HCI research (and other related fields such as psychology), we are often trying to understand people and their interactional context. This represents a shift from utilizing the time and skill of the users, as found in CitSci, to studying the users themselves, as found in Psychology and HCI research.

The kinds of tasks that HCI researchers are interested in also differ from typical crowd sourced tasks. HCI experiments often require participants to pay attention for extended periods of time, without deviation. Response time, for example, is often a very important metric used in experiments. Using such a metric requires participants to concentrate and react as quickly as they can to stimuli. Crowdsourcing tasks such as the classification of images do not require users to pay close attention for long periods of time, nor do they require that the task be completed as quickly as it can, it is often possible for the citizen scientist to break from a task whenever they wish without affecting data quality. The collection of QS data is even further removed from the strict procedures of controlled experimentation as it is often collected passively over extended periods of time.

If we take the area of typing studies as an example, we can see why one might expect HCI experiments conducted with citizen science volunteers or those with an interest in QS to produce incomplete and unreliable data due to their very different methodologies. Indeed, that is assuming that participants could be recruited at all. Participation in experiments is usually reimbursed with cash or vouchers because there is often no gain to the participant to volunteer for a time consuming and boring activity. It is entirely possible that no one would wish to take part in such work without reimbursement, but as there have been relatively few previous attempts to use citizen scientists in HCI research this is an open question.

In this paper we present a report of our attempts to run a 2 (location) x 3 (reimbursement) condition typing study conducted in order to test the effectiveness of alternative reimbursement methods in HCI research. The study is run in different locations (lab and online) and with different remuneration levels (money, data, none) to reflect standard recruitment methods, and Citizen Science and Quantified Self data collection methods respectively. We explore how participation levels are affected by location and remuneration type, and how these conditions affect data quality when compared to traditional lab-gathered data. We show that it is possible to recruit participants successfully without offering monetary reimbursement, by offering participants both data about their typing performance, and simply calling for aid in scientific research. The experiment required a higher level of engagement from participants than the more passive forms of data collection associated with Quantified Self. We additionally show that in

terms of data quality, citizen scientists produce data similar to that gathered in the lab, and that data gathered online from *paid* participants may actually be less reliable in terms of replicating effects found in the lab.

2. BACKGROUND

2.1 Lab Experiments in HCI

The aim of lab-based experiments in HCI is to isolate and study a particular aspect of human behavior when interacting with technology. Studying such interaction "in the wild" offers a chance for researchers to understand the true use of a particular technology in the context of the real world (Rogers, 2011). However, the issue with such an approach is that it is impossible to understand fully the vast array of elements affecting the interactions they are studying; with a large number of confounding factors it can be hard to pin point causality between the technology and behavior. By moving these interactions into the lab with a designated task to complete, a wide array of elements can be controlled, meaning the researcher is able to manipulate solely the aspect they wish to study, whilst keeping other features fixed. This makes controlled lab experiments a useful and commonly used tool within HCI research.

Controlled HCI experiments often require participants to complete small micro-tasks over a number of repetitions with the aim of understanding how variations in these tasks can affect the participant's performance. Naturally, these tasks can vary a great deal: research relating to typing for example will often involve participants copying tens or hundreds of sentences on the whichever device is being studied (Dunlop, Komninos, & Durga, 2014; Dunlop & Levine, 2012; MacKenzie & Soukoreff, 2003; Salthouse, 1986) whereas research aiming to understand how users multitask when using various technology might ask participants to follow a set of instructions in order to interact with a specially designed interface (Brumby, Cox, Back, & Gould, 2013). To understand the effect these variations have upon the interaction, a number of metrics are recorded during these tasks, from the quantitative speed and accuracy readings, to more qualitative user feedback and ratings. A combination of these results can be statistically analysed to then determine whether the researcher-controlled variations within the task significantly affected the user interaction with the device or task being studied.

2.2 Experiments Online

The repetitive tasks and easily recorded metrics used in HCI experiments make them a prime candidate for moving to an online platform, thus avoiding the need for participants to travel to a lab location saving time and money for both researcher and participant. Indeed, an increasing amount of HCI research is beginning to make use of the Amazon Mechanical Turk (AMT) crowdsourcing platform ("Amazon Mechanical Turk," 2015). AMT matches "workers" to researchers and developers, who produce online software to be used and tested by AMT workers. The workers are reimbursed for their time, which they can allocate whenever they wish, and the

researchers get data. One key difference between online- and lab-based experiments is the lack of oversight that researchers have over the participants' work; without the researchers to observe the task being completed, the participants may misunderstand the task or complete it without care and attention thus affecting the results. For this reason, much work has gone into understanding more about the quality of data gathered using this manner (Buhrmester, Kwang, & Gosling, 2011; Crump, McDonnell, & Gureckis, 2013; Gould, Cox, Brumby, & Wiseman, 2015; Kittur, Chi, & Suh, 2008). Komarov, Reinecke, & Gajos, (2013) show that the data collected in a lab and the data collected from AMT can be highly comparable. Gould et al (2015) show that it is possible to collect high quality data even on longer tasks that require a high level of concentration. These results suggest that HCI research conducted using online platforms can still produce valid results. However, this finding is not replicated in all online experiments, as noted in Ramsey, Thompson, Mckenzie, & Rosenbaum, (2016), success rates when replicating lab experiments online are not consistent.

Despite the fact that recruiting participants via AMT can be substantially cheaper than doing so by other means, questions have been raised regarding the ethics of reimbursing AMT workers at a lower rate than would be done in a lab (Crump et al., 2013; Mason & Suri, 2012). With this in mind, there is a belief that experiments conducted on AMT workers should be regarded in a similar way to those in controlled laboratory experiments (Crump et al., 2013) and that the participants should be reimbursed accordingly. This would mean that although using AMT for recruitment can make running controlled HCI experiments faster, it should not necessarily result in smaller reimbursement costs.

Other studies suggest that reimbursement rate may have an effect upon the quality of data that can be gathered by paying on online. Mason & Watts (2010) offered varying levels of reimbursement and found that participants completed more work if paid more, but the quality is not improved by offering increasing sums. However, they did not investigate the effect of no monetary reimbursement.

2.3 Citizen Science

Whilst there is concern regarding the ethics of differing payment levels for lab-based and web-based participants, it is possible to encourage participant engagement with an online project with no monetary reimbursement at all. The relatively new branch of science that is Citizen Science taps into the public's desire to help in scientific research, and there are many projects making use of this crowd sourcing (see (Bonney, Cooper, et al., 2009) for a review). The tasks that volunteers are given to complete often involve using them as data gatherers, or processors, rather than as the data source.

When considering how citizen science might be applied to HCI, it is useful to understand what motivates people to offer their time to these projects for no monetary reimbursement and research

has shown a diverse set of motivations . Galaxy Zoo, a curated website containing a number of online Citizen Research topics conducted research to find out the motivations for taking part. This research suggested that a variety of motivating factors brought people to the site, including a desire to contribute to research, an interest in the topic of astronomy and the incentive of looking at galaxies that no one else had seen before (Raddick et al., 2010, 2013). The importance of interest in the topic has been seen elsewhere, for example, in the research conducted by Evans, Abrams, & Reitsma (2005) which relied upon the citizen scientists' inherent interest in their local area when collecting data about birds in the environment nearby. Other studies have shown that reasons for engagement in such studies can change over the course of a study, and are affected by the levels of interaction between researchers and citizen scientists (Rotman et al., 2012). All agree that the sense of contribution is important along with an interest in the subject. It seems possible that HCI experiments may be able to make use of this sense of contribution, but the need for an interest in the subject may prove harder. Previously successful projects have involved engaging participants in novel subjects, or using engaging, game-like presentations (Von Ahn & Dabbish, 2008). Often HCI research focuses on less engaging subjects, this may mean that incentivizing participation in CitSci studies in HCI is difficult. Any HCI experiments that use citizen science recruitment methods may need to appeal to groups of people who are already interested in the factors that influence how people use computers.

It is also suggested that a sense of community is a motivator for participation in citizen science research (Jennett, Eveleigh, Mathieu, Ajani, & Cox, 2013). On many CitSci sites, participants are encouraged to return time after time and continue completing tasks. The occurrence of repeat visitors means that strong communities can emerge, chatting on forums and discussing the tasks on the site. This sense of community has been shown to be a strong motivator for participants (Jennett et al., 2013). Again we find a potential problem when applying this motivator to HCI research. Controlled HCI experiments often involve some level of deception to ensure that the participants act as naturally as possible, without adapting their behavior to fit the experimenter's hypothesis. This arrangement requires that participants specifically do not talk to one another, nor take part in the experiment again. For this reason, communities are less likely to be a motivating factor for citizen scientists taking part in controlled HCI experiments.

2.4 Quantified Self

Unlike the citizen science approach, the personal informatics, or "Quantified Self" movement is concerned with gathering data about the self. This differs from personal data collected for CitSci purposes, as the data will only be viewed and analysed by the owner of the data, not by external researchers (this is a potential barrier to participation in CitSci projects (Jennett et al., 2014)). This field is interested in the way that people log data about themselves, from simple fitness apps, to taking minute by minute photographs (see (Li, Dey, & Forlizzi, 2010; Swan, 2012) for reviews). People are becoming more interested in gathering data about themselves in order to find out more about the way they act in the world and allow for more informed reflection. Munson, (2012) suggests there may be two different approaches to QS: persuasive and mindful. Some

people may choose monitor data about themselves in order to remain inspired to change something (for instance trying to complete 10,000 steps a day) whereas others may take part in order to better understand themselves (for example understanding their sleep patterns). Other researchers have suggested that simply having access to numerical data about oneself is a strong motivator (Fritz, Huang, Murphy, & Zimmermann, 2014).

This form of participant incentive has been used in HCI research, see the Lab In The Wild ("Lab In The Wild," 2015) and Test My Brain ("Test My Brain," 2015) projects. These sites host a series of experiments online in which participants can take part, in exchange for finding out new information about themselves. On these sites, the experiments are framed specifically in terms of what the user can learn about themselves, on some the user is offered the opportunity to see how their responses compare to others who have completed the tests previously. Although the word "citizen scientist" is sometimes used on these sites, the tasks area always presented with some form of data-based reward. The Test My Brain site specifically states that "the study should be fun for our participants" if any researchers wish to submit their own experiments. An investigation into participant motivation shows participants take part because they find the studies fun and enjoy comparing their results to others (Reinecke, Arbor, & Gajos, 2015). It appears that enjoyment is key for participation in online experiments such as these.

The studies described above have been successful in recruiting participants: the data collected from these websites has fed into various publications (see (Germine et al., 2012; Halberda, Ly, Wilmer, Naiman, & Germine, 2012)), suggesting the quality of data collected using these method is high. However, these sites rely on being able to provide participants with interesting and novel data about themselves; it may not be the case that all HCI experiments lend themselves to producing such data. Can participants be satisfactorily incentivized from simply having numbers to describe seemingly mundane aspects of themselves that are of interest to HCI researchers?

2.5 Comparison of methods

It is important to understand how these three methods (standard reimbursement, CitSci and QS) compare to one another in terms of participation levels and data quality. All three have produced peer-reviewed research but it is not clear how these approaches affected the data produced. Is data quality affected when participants are motivated altruistically compared to monetarily? Is it more difficult to recruit participants who are not given any data or monetary reimbursement? Without a direct comparison of these methods using the same experimental paradigm, it is difficult to know how these recruitment methods compare to one another.

Recruitment (finding out about and taking an interest in a project) and participation (choosing to commit time to a project) are subtly different stages in running an online study, both with a range of different motivators, from interest, ease of entry and sincerity (Organisciak & Twidale, 2015).

In this paper we aim to explore solely the effect of reimbursement and not other possible motivators.

In this research we tie together the alternate forms of participant motivation discussed above and explore their effect on the participation levels and data quality of a simple typing study. We show that despite a lack of monetary incentive, participants can still be found for HCI experiments, despite their lack of apparent novelty. Drop out rates are higher when participants are not paid, but initial uptake is high enough for this not to negatively affect participation rates. We show that high quality data can be gained from both CitSci and QS methods. Curiously, it appears that data quality may be lower when participants are paid, but not observed in the lab, suggesting that accountability has an effect upon the effort that participants put into an experiment.

3. METHOD

We ran a typing experiment with six different recruitment conditions. The typing experiment did not vary between these conditions so that the results could be compared between recruitment conditions. One of the six conditions acted as a baseline; this condition replicated the standard recruitment practices often used within HCI research as participants were asked to come to a lab and were reimbursed for their time. Data quality from any other recruitment methods will need to match that gathered in the lab to be considered acceptable.

3.1 Design

The experiment was a 3x2 between participants design. The two independent variables were incentive and location. Location had two levels: **Lab** and **Online** whilst incentive had three levels: **Citizen Science** (no reimbursement), **Quantified Self** (reimbursement in personal data) and **Reimbursement** (monetary payment for participation). The base line was represented by the **Lab** location with **Reimbursement** as an incentive. These were between participant conditions, with each participant taking part in only one of the six conditions.

The dependent variables being measured were levels of participation and data quality. Participation levels were judged not only on how many people completed the experiment, but how many completed the experiment given the total number of those that started it. This measurement would provide a drop-out rate. Data quality was assessed based upon measurements taken during the study. The standard for HCI controlled experiments is data collected from lab-based studies with paid participants and therefore this would act as a benchmark when assessing data quality. Data would be considered good quality if it reproduced results in the same direction with a similar level of significance as those collected in the lab reimbursed condition.

3.2 **Participants**

A total of 117 participants completed the experiment; 44 females, 64 males and 9 undisclosed/other. The average age of participants was 25 years . The total of 117 does not include participants who started the experiment but did not finish; demographic data about these participants could not be collected.

Participation rates were a key dependent variable measured in this experiment, for that reason further details will be given in section 4, *Results*. Recruitment method details are described in section 3.4, *Procedure*.

3.3 **Materials**

The experiment used a typing "game", which involved participants transcribing text presented on screen. This was a replication of the studies reported in Salthouse, (1986), which aim to document multiple different typing "phenomena". The phenomena cover metrics from typing speed, to error rates, to multitasking ability whilst transcribing, all of which can be predicted to a certain level of accuracy for typists given their experience and age. In this experiment, we specifically tested the 'Eye Hand span' of our participants. This span is the number of characters of the text that a typist must be able to see in order to type at optimal speed. The span is tested by altering the preview size (number of characters visible) of the text being copied. Previous research has shown that the smaller the preview size, the slower users are able to type compared to their normal speeds (Salthouse & Saults, 1987). The point at which users are able to type at their optimal speed is termed their "Eye Hand Span". On average, typists require a preview of 5 characters to type optimally (see Figure 1 for an explanation). Of particular interest is the fact that a typist's Eye Hand span reduces as the material that they are copying tends to randomness – the Eye Hand span is shorter for non-words than it is for words. In this study, the aim was to understand how the Eye Hand span is affected when copying numbers. Thus in the game, participants were asked to transcribe real words, randomly generated strings and numbers. These were presented in counterbalanced orders to the participants. Eight different preview window sizes were used throughout the experiment, these too were counterbalanced for order.

This typing game was used for this particular investigation into recruitment and incentive in HCI studies because it can be easily deployed online, and is a relatively simple task for users to complete without the guidance of an experimenter. It also does not require large amounts of processing power and therefore minimizes the number of potential participants who might be excluded due to lower specification home computers that may be used to complete the task.

Current key stroke   Preview windows of 13 characters

The quick brown fox jumped over the lazy...
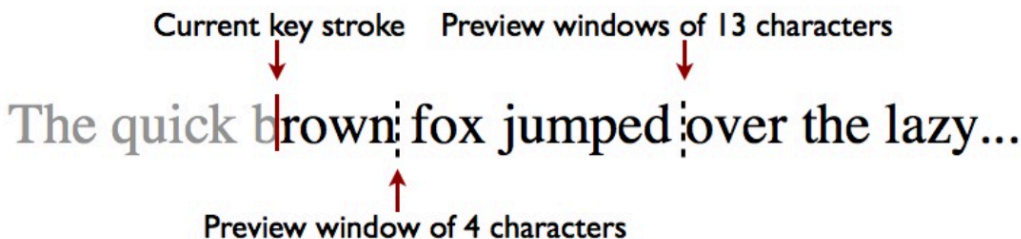
Preview window of 4 characters

Figure 1: Illustration of the eye-hand span. If the preview window is reduced to 4 characters (in this situation the typist would not be able to see past the word "brown"), which is smaller than the desired 5 character Eye Hand span, the typist will slow down and will not be able to type at optimal speed. With a larger preview window of 13 characters (the typist could now see the word "jumped") the typist can see further ahead and can therefore type at optimal speed.

The experiment was presented as a website. The first page gave participants information on what the study would involve. In the Reimbursed and QS conditions, participants were informed that they would be able to see their performance data at the end of completing the task for both words and numbers (note, this was not mentioned in the study advert for the Reimbursed conditions so was not an incentive). Participants were warned against participating if suffering from RSI. At this point participants chose to take part in the experiment or leave. If the participant chose to go on, the second page provided more detailed information and included a video showing a sample game, which paused and highlighted key concepts of the game that were needed in order to progress. After this, participants were given a trial period (of 10 words, or numbers dependent upon their in-experiment condition) to play the game. They were then presented with the game itself.

During the typing game, text was presented to the participants (either words or numbers) at the top of a black screen in white font. Participants were instructed to copy this text; their copied text could be seen in the top left corner in red. To make the game visually engaging, coloured orbs appeared on screen. The more words/numbers the participant typed correctly, the more orbs appeared. Incorrect words/numbers resulted in half of the orbs disappearing. See Figure 2 for a screenshot of the game.
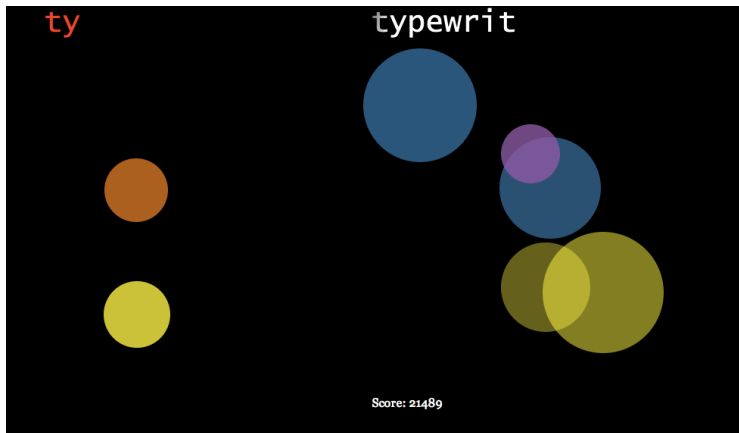
Figure 2. Screenshot of the game when running. The target text ("typewrit") can be seen in white at the top of the page. The typed text ("ty") can be seen in the top left in red. Coloured orbs can be seen, showing that the participant has previously successfully typed a number of words. The score is shown at the bottom of the screen.

The 320 words/numbers that the participants typed during the game displayed as a continuous stream of text. The number of characters in the stream that were visible to the participant varied throughout the experiment between 1 and 8 characters per block of 10 words/numbers. This often meant that the end of a word/number could not be seen (see Figure 1, the end of the word "typewriter" cannot be seen). The stream of text that the participant copied progressed leftwards by one character each time the participant typed a character, a method used in previous incarnations of this experiment.

A scoring system was implemented to encourage participants to manage their speed and accuracy. More points were given for a speedily entered word/number, and points were lost for submitting an incorrect word/number. Participants were informed prior to the game of the how the scoring mechanism worked. Players were able to use the backspace key to correct any errors they noticed before submitting the word/number. However, after pressing the space or enter key they could not go back to correct a word.

Data about performance was available to participants in the QS and Reimbursed conditions once participants had completed both the word and number in-experiment conditions. The results page displayed an explanation of the study, the participant's calculated Eye-hand span (Salthouse, 1986), a table displaying the mean, minimum and maximum typing time per character in the 8 different preview sizes, and 4 box graphs: 2 interactive Google charts ("Google Charts," 2015) and 2 static, downloadable graphs showing their performance over the 8 different preview sizes (see Figure 3). The participant's score was also displayed alongside the average score and the number of people who had played the game. The participant was informed if they had beaten the high score for that particular game. The participants were also given a link to allow them to download their raw data in the form of a 'csv' file. This file showed timing data for each keypress made throughout the experiment, and highlighted correct and incorrect entries.
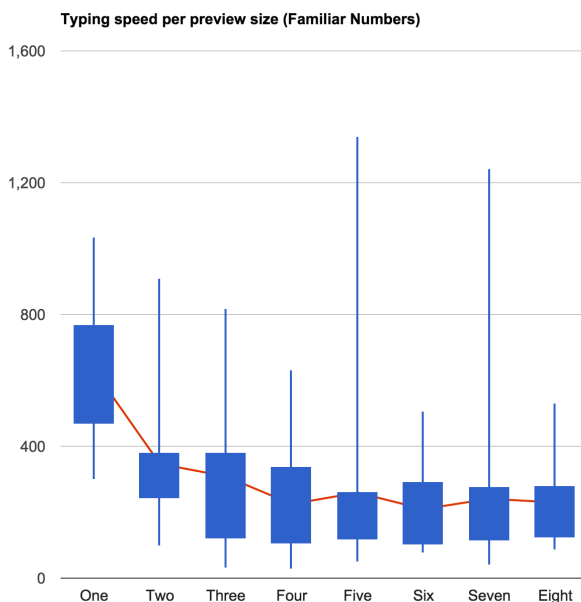

Figure 3. Example chart provided to participants in the Reimbursed and QS conditions. The x axis represents the preview size and the y axis displays average time in milliseconds to type a word/number. In this example, the participant's speed increases as the preview window gets larger. Participants were able to reflect upon their typing performance using the charts provided.

3.4 Procedure

Baseline results for the experiment were required in order to provide a benchmark with which to compare results from future conditions. The first condition therefore was the **Lab Reimbursed** condition. An advert was placed on a University Subject Pool. This pool was open to university students, staff and members of the local community. The advert stated that participants would come to the lab to complete a game-like experiment and would be reimbursed ~USD$4.20 for taking part in a 20-minute experiment. Once participants arrived in the lab environment, the experimenter sat them at the computer from which they read all instructions and watched the video. The experimenter did not offer additional guidance unless it was requested in order to better simulate the online condition, where participants would have to complete the experiment using only the information provided to them on screen.

Once all participants had been recruited and the results had been collected, the advert was taken down. A new advert was placed for the **Lab QS** condition using the same University Subject Pool. As participants would have to travel to the lab, it was important that this experiment was advertised in a local area, thus the subject pool was chosen rather than a specific QS venue. The advert asked participants to come to the lab to take part in a game-like experiment, which would tell them more about the way they typed. No monetary reimbursement was offered. This advert was left up for one week before it was taken down. The final lab-based advert for the **Lab CitSci** condition was placed on the same University Subject Pool for the same reason of participants needing to be able to easily travel to the lab. Participants were asked to take part in a game-like experiment in order to aid a research project. Participants were offered no money, or data in return for their participation. This advert was taken down after one week.

The **Online Reimbursed** condition was run next. Our intention was to replicate the Lab Reimbursed condition with the exception of location in which the study was conducted. Therefore this experiment was advertised on the same University Subject Pool. A setting was used that ensured no one who had taken part in a previous study could reapply. Participants were offered ~USD$4.20 in return for taking part in the 20-minute online experiment.

The **Online QS** condition was advertised on the aggregation website *www.reddit.com* in two specialist 'subreddits' (a group that focuses on one topic): the Quantified Self and Cognitive Science subreddits. The advert asked users to participate in a typing game that would provide them with information about the way they typed. Results were collected for a week after posting, though people could continue to participate after this time. The study was not advertised in places (such as social media) that would influence people to partake purely because they knew the experimenter. This location was chosen for the advert, as opposed to the University Subject Pool, as we were aware recruitment would be more successful with people who were interested in QS. This would also be more representative of the likely participant population of any QS based HCI experiments in the future.

The **Online CitSci** condition was advertised again on the *www.reddit.com* website. A new account was used to post the study and it was advertised in a different subreddit (the experimental psychology subreddit) so as to avoid people seeing the advertisement for the previous experiment in quick succession in the same location, although it is possible that people subscribed to both subreddits might have seen both adverts. The advert asked for participants to take part in an online study and did not specify any reimbursement in any form. Again the results were collected for one week.

4. RESULTS

This section will report the results of the experiment into recruitment methods. This will cover **Participation Levels** and **Data Quality**. Participation levels are assessed both in terms of how many participants completed the experiment, and how many dropped out midway through. Data Quality is assessed as whether the results of the experiment echoed those of the base level results recorded in the Lab Reimbursed experiment: comparisons between speed and error rates will be made. The results of the typing experiment will be briefly reported in order to compare between conditions but will not be discussed in this paper.

4.1 Participation Levels

Table 1 summarizes the final participation levels for each condition including demographic details. This table represents only completed trials with usable data. Any participants who took longer than 500ms to press a key were removed from the data as this suggested a lack of attention paid to the experiment. Recruitment for the Lab Reimbursed condition was the only successful lab-based condition, for the Lab CitSci and QS conditions, no participants signed up for the study.

Condition	n	Female/Male/Other	Mean Age (SD)
Lab Reimbursed	14	10/4/0	29.4 (8.1)
Lab CitSci	0	n/a	n/a
Lab QS	0	n/a	n/a
Online Reimbursed	14	9/3/2	24.4 (9.8)
Online CitSci	44	16/26/2	24 (6.7)
Online QS	45	9/31/5	24.5 (10.6)

Table 1: Participant demographics in each of the six experimental conditions.

Table 2 summarizes the participation levels throughout the experiment. A complete experiment is defined as completion of *both* the words and numbers conditions. This highlights the differences in drop-out rates between the conditions. In the Lab Reimbursed condition, no participants dropped out of the study. Two participants' data were removed due to long key presses. The drop out rate was also 0% for the Online Reimbursed condition. Five participants data were removed

due to long keypresses. One participant's data were removed as they did not understand the experiment.

Condition	Signed up	Started Experiment	Finished Experiment	Dropout rate
Lab Reimbursed	16	16	16	0%
Lab CitSci	0	0	0	n/a
Lab QS	0	0	0	n/a
Online Reimbursed	20	20	20	0%
Online CitSci	395	322	47	88.1%
Online QS	173	158	44	74.6%

Table 2: Participation levels in each condition. Note those who finished the experiment did not necessarily produce useable data.

Recruitment for the Online QS condition was very successful. The total number of completed single trials (trials where a participant only completed either the words or numbers in-experiment condition) was 123 for numbers and 152 for words. Some participants started the experiment but did not complete a single condition. The total number of completed full experiments (the participant had completed the game for both words and numbers) was 47. Two participants' data were removed for having key presses longer than 500ms. If a drop out in this experiment is defined as either aborted trials, or only completion of a single trial, this puts the drop out rate at 88.1% as only 47 of the 395 participants who signed up to the study completed it fully.

Participation levels for the Online CitSci condition were also high. The number of completed single trials in the study was 48 for numbers and 66 for words. The total number of complete experiments was 44. Defining drop out rate in the same way as in the Online Quantified Self experiment, this study had a dropout rate of 74.57%. In the Online QS condition, no data was removed.

4.2 **Data Quality**

Standard controlled experiments within the domain of HCI typically take place in a lab setting with participants being given reimbursement for their time. In the present experiment, the Lab Reimbursed condition best reflects this typical set-up. In order to determine that the alternative recruitment methods tested in this current experiment are suitable, they would need to reflect the results gather in the Lab Reimbursed condition. Note that no Data Quality analysis can be performed on the Lab QS and CitSci conditions as no data was collected.

Here we report the results of the Lab Reimbursed experiment as a form of base line. The further experiments will be compared to the results presented here. The analyzed data comprised only

words and numbers that were typed correctly; trials containing errors were ignored for timing analysis.

The median length of time taken per character for numbers was 352 ms, and for words was 331 ms. A repeated measures ANOVA reveals that this difference was significant, $F(1,12)= 4.9474$, $p<.05$: words were faster to transcribe than numbers.

A repeated measures ANOVA showed that the effect of preview size was also found to be significant, $F(7,84)= 167.69$, $p<.001$: the larger the preview size, the faster participants were able to type. These tests would be replicated for each of the following analyses. A summary of all results can be found in Table 3.

In the Online Reimbursed condition the median keypress length per character for numbers was 388 ms. The average keypress length per character for words was 348 ms. Statistical analysis shows this difference to be non-significant, $F(1,12)= 2.2096$, $p=.1629$. This does not replicate the results from the Lab Reimbursed condition. The effect of preview size however, was found to be significant, $F(7,84)= 83.034$, $p<.001$, which replicated the finding in the lab reimbursed condition.

In the Online QS condition the average speed per key press per character for numbers was 315 ms and for words was 252 ms. These are both faster than the baseline established in the Lab Reimbursed condition. However, both of the significant effects in the baseline condition were replicated in this condition: there was a significant effect of target type (words or numbers) on typing speed, $F(1,38)= 31.346$, $p<.001$, and the preview size also had a significant effect, $F(7,266)= 296.37$, $p<.001$.

In the Online CitSci condition numbers took on average 326 ms per key press and words took on average 250 ms, this was again faster than found in the Lab Reimbursed condition. There were two main effects on typing speed found in the study: that of target type, $F(1,40)= 49.83$, $p<.001$, and preview size, $F(7,280)= 491.5$, $p<.001$. This replicates the results of the baseline.

5. DISCUSSION

This study aimed to investigate how alternative recruitment techniques that used non-monetary incentives such as QS and CitSci affected recruitment and data quality of controlled HCI experiments. We hypothesized that participation levels may not be high enough due to the lack of monetary reimbursement, the lack of novelty in the experiment and the mundaneness of the data given to participants at the end. For these reasons, it was also possible that the quality of data collected would be reduced. These hypotheses were, however, unsupported by the data.

Condition	Median time per char – Words (ms)	Median time per char – Numbers (ms)	Significant difference between Words and Numbers	Significant effect of Preview Size
Lab Reimbursed	331	352	YES	YES
Online Reimbursed	348	388	NO	YES
Online CitSci	250	326	YES	YES
Online QS	252	315	YES	YES

Table 3: Summary of results for each condition. Lab Reimbursed represents the "base line". The Online Reimbursed condition does not replicate both significant effects whereas the other Online conditions do.

5.1 Participation Levels

Participation levels for the non-remunerated lab-based conditions were zero. This is not to suggest that it is impossible to recruit in this manner when the participants need to travel (rather than completing a task online), but it does suggest that recruitment using this method is more difficult. The population in the University Participant pool are used to being paid for their time so there is little incentive to taking part in an experiment for free. Finding a place to advertise the study where locality and interest overlap may be key. Advertising the study simply to people in the local area (as done in this paper) may not yield results. Advertising to local *interest* groups in the local area however may be more successful as previous research suggested that a key motivator for people to take part in non-reimbursed scientific experiments is an interest in the topic itself (Raddick et al., 2013). The results of this experiment suggest more care needs to be taken when advertising lab-based CitSci or QS recruited HCI experiments.

These studies did however, show that people are willing take part in *online* experiments, even if they do not get reimbursed monetarily. The success of previous initiatives aimed at recruiting people to experiments using data as a reward suggested that this would be possible (Reinecke et al., 2015) however, the current study shows that this is also possible even without a carefully designed "fun" sounding experiment. This suggests that QS-style recruitment might be possible for a wide range of experiments, not just those that will provide users with particularly novel data about themselves. Simply sharing the data collected from the user's experimental trial appears to be enough of an incentive. Although currently it's not clear if the raw data would be enough to encourage participation, or if the brief analysis and presentation of the data in graphical form as used in this study was necessary or helpful.

The current study also showed that people were willing to take part in this experiment without receiving any clear form of reimbursement, data or otherwise. This has been seen many times in the Citizen Science community, where people who do not gain monetarily, still take part in online research for intrinsic reasons, often due to interest in the subject alone (Raddick et al., 2013). However, this is often rewarded with inclusion in an ongoing online community (Jennett et al., 2013) which was not possible in the current study, which required a one-time interaction with no conversation between citizen scientists. This result suggests that single interactions, rather than prolonged, multiple interactions, are a small enough time commitment that no other rewards are needed.

The number of completed full trials in both the online QS and online CitSci experiments was very similar. However, both experiments suffered from high dropout rates, which is a common problem in online experimentation (Dandurand, Shultz, & Onishi, 2008). The dropout rates were higher in the QS experiment (88.1%) than they were in the CitSci experiment (74.57%) and the total number starting the QS experiment was over twice that starting the CitSci experiment (395 v 173 respectively). This suggests that those in the two conditions may have been motivated for different reasons. Participants in the QS condition may have been motivated solely by finding out information about themselves, rather than helping with the research project itself: once they had received some data, they were happy to quit before completion. Those in the CitSci condition appeared less likely to do this, perhaps because their goal upon beginning the experiment was to complete it. This study suggests that the QS approach of offering participants data about themselves attracts a large number of potential participants, but the CitSci approach may ensure higher completion rates.

These studies show that reimbursement is not necessarily required for participant recruitment *if the study is run online.* In order to encourage people to come to the lab however, it seems that an incentive is required. Final participation levels in the online CitSci and QS conditions were comparable, however the dropout rate was not. In this study it appears that participants were motivated strongly by the opportunity to obtain data about themselves, and once they had that data they were more likely to leave the experiment. CitSci incentives however appear more likely to encourage completion of the experiment.

5.2 Data Quality

The expected results of this experiment were that preview size would have an effect upon typing speed and would make **typing slower as preview size decreased** (as was seen in (Salthouse, 1986)) and that words and numbers would be typed at different speeds, based upon previous research showing that randomized text is slower to type than predictable, word-based text suggesting **numbers would be typed slower than words** (Salthouse, 1986).

The basis of this study was that the Lab Reimbursed condition would act as a baseline, as it best reflects the standard conditions of controlled experiments within HCI. In the current study, the

Lab Reimbursed condition did confirm the hypotheses that words were faster to type than numbers and smaller preview sizes resulted in slower typing. We might therefore expect this finding to replicate regardless of the form of participant recruitment technique. Any deviations may suggest an issue with the recruitment method interfering with the experimental results.

In all conditions with participants, the effect of preview size on typing speed was found to be significant and thus replicated the baseline condition. However, the hypothesis that words would be typed faster than numbers that was found in the Lab Reimbursed, and replicated in the Online QS and Online CitSci experiments, was *not* replicated in the Online Reimbursed experiment. Whereas words were significantly quicker to type than numbers for most participants, this was not the case for those in the Online Reimbursed condition. The replication of this finding in other online conditions, suggests that this effect should be found consistently. This effect is also found in previous literature showing that when strings become random and unpredictable (as numbers are) users type slower (see (Salthouse, 1986) for a review). This again suggests that we might expect numbers to be transcribed slower than words. A deviation from this may represent poor data quality. Participants in the Online Reimbursed condition also had the slowest typing speeds for both words and numbers and had five participants removed due to long key presses which may suggest that they were not as engaged with the task as those in other conditions. Additionally one participant clearly did not understand the instructions for the experiment. This suggests they may not have read the instructions thoroughly, again suggesting that engagement with this study was low in the Online Reimbursed condition.

That there was a failure to replicate the difference in typing speeds between words and numbers, and slower typing speeds in general in the Online Reimbursed condition may highlight a potential downside of monetarily motivated online participants, whose main aim is to complete the experiment and receive payment. If the payment is not related to the quality of data that is produced, there is no incentive to engage with the experiment – they will be paid regardless – a phenomenon noted by Mason & Watts, (2010). Whereas in the lab, participants are incentivized to complete the experiment to the best of their ability due to researcher presence perhaps or the "scientific aura" (Ramsey et al., 2016), online there is no such incentive meaning those in Online Reimbursed conditions can choose to complete the task without giving it their full attention.

Those people who took part in the online QS and CitSci conditions however, were motivated to take part for other reasons and therefore had a stake in the experiment being completed properly in order to generate accurate data. Participants in the QS study for example needed to complete the experiment correctly in order to find out information about themselves. Any distraction from the study would have resulted in them receiving skewed and false information, which goes against their motivations for taking part in the first place. Those in the CitSci study were likely to be motivated by simply helping out in scientific research, and would therefore care about the quality of information they were providing (as in (Raddick et al., 2010)).

Yet despite the fact that in this study, the baseline was not replicated with the Online Reimbursed condition, many people have reported successes in this area (Dandurand et al., 2008; Heer & Bostock, 2010; Kittur et al., 2008; Komarov et al., 2013). The study presented here is different in some ways to these papers however. For one, our study used participants solely recruited from a university subject pool paid with money, not course credit, whereas most studies use AMT workers. There could be differences in motivation and desire to complete work to a high standard between these two groups. AMT workers require high ratings in order to get more work. For the potentially one-off participants recruited in the current study, there was no such necessity. Further research will need to establish whether participation rates affect the quality of data that participants produce: do those who complete experiments for a living perform better than those who do it as a one-off online?

Despite the main effects found in the Lab Reimbursed condition being replicated in the online QS and CitSci conditions, there do appear to be differences in the speed of typing. Participants in the QS and CitSci experiments typed faster than those who were reimbursed in both the lab and online. This need not be an issue in terms of interpreting the results, as this was not a test of overall typing speed, but about the interactions between target type and preview size and base typing speed. What is does show is that those who took part in this study voluntarily were more familiar and at ease on a computer than those who were paid. Firstly, they were recruited from an online forum and so were likely to be highly computer literate. Secondly, the sample of people who put themselves forward to take part in these studies are likely to be biased to those who are already good at the task being tested in the experiment, in this case, good typists. Ultimately this did not matter for the results of this experiment, as the preview size effect has been replicated between both fast and slow typists (Salthouse, 1986). However, this does highlight a particular issue with investigating HCI topics online; by sampling from an online population, you will naturally collect computer literate participants. This need not be a problem in other research areas with focuses on topics other than computing (for instance pure psychology experiments) however it certainly needs to be considered when reporting online HCI experiments.

5.3 **Limitations**

One limitation of our study is that it is difficult to ensure that the online QS and CitSci experiments were advertised to the same *sorts* of people. It was not possible to advertise the study twice in the same area, as participants who had taken part in the first condition would not want to take part again. Therefore, similar, but not identical subreddits were chosen. However, the advertising strategy aimed to recruit different types of people; for instance we wanted those who were interested in QS methods to engage in the Online QS condition, whereas we wanted those interested in HCI-style experiments to engage in the Online CitSci condition. This decision was made in order to better reflect the types of people who would naturally be attracted to the studies, depending on how they were advertised. It is likely that any online study offering data as a repayment would naturally attract potential participants with an interest in QS.

To generalize this approach to more HCI experiments, it will be important to identify key communities to advertise future experiments. Although existing sites exist (for example Zooniverse), these may not be an option for all researchers who are unable to produce online experiments at a high enough standard or with a novel enough "hook". Advertising on reddit worked for this particular experiment, but this is not a sustainable approach for future research; flooding communities with studies may alienate people from taking part. This is a limitation that will have to be overcome as HCI research progresses in this area.

One final limitation that we wish to address is the possibility that the lack of replication found in the Online Reimbursed condition may be an anomaly, and in other cases it may replicate. We acknowledge that further research into this will be necessary to make stronger conclusions about this effect, although previous research has suggested differences in performance between lab-based and online participants (Dandurand et al., 2008).

6. CONCLUSION

The success of the Quantified Self and Citizen Science recruitment in other fields suggests that such approaches may be beneficial for HCI research. If successful, this would make HCI research cheaper, and more efficient. In this paper we conducted a study to compare three different recruitment methods (reimbursement, quantified self and citizen science) at two different locations (online and at a university lab) on participation levels and data quality in a typing study.

We showed that these new incentives can indeed be successfully used to recruit for online HCI studies, and that the participation levels and data quality are high, replicating results found in the lab. However, questions are raised about how participant motivation can affect participation levels and data quality. Paid online participants may not be invested in the research, and may therefore produce lower quality data, as in this study this condition did not replicate the lab results. Although QS and CitSci do motivate people to take part, participants incentivized by the quantified self may exit the study once they receive data about themselves.

Previous research suggests that to run successful studies online, participants will need an interest in the topic, a community and for the study itself to be fun and novel. However, we have shown in this study that it is possible to run HCI experiments online without providing community and novelty.

This study shows that new methods of recruitment may benefit the HCI community. There appears to be little difference in final participation levels and data quality, suggesting that both QS and CitSci recruitment methods are viable for future HCI and other online experiments.

7. **REFERENCES**

Amazon Mechanical Turk. (2015). Retrieved from https://www.mturk.com

Bertamini, M., & Munafo, M. R. (2012). Bite-size science and its undesired side effects. *Perspectives on Psychological Science*, *7*(1), 67–71. http://doi.org/10.1177/1745691611429353

Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J., & Wilderman, C. C. (2009). Public participation in scientific research: defining the field and assessing its potential for informal science education. *A CAISE Inquiry Group Report*.

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience*, *59*(11), 977–984. http://doi.org/10.1525/bio.2009.59.11.9

Bravata, D. M., Smith-Spangler, C., Sundaram, V., Gienger, A. L., Lin, N., Lewis, R., … Sirard, J. R. (2007). Using pedometers to increase physical activity and improve health: a systematic review. *Jama*, *298*(19), 2296–2304. http://doi.org/10.1001/jama.298.19.2296

Brumby, D. P., Cox, A. L., Back, J., & Gould, S. J. J. (2013). Recovering from an interruption: Investigating speed   accuracy trade-offs in task resumption behavior. *Journal of Experimental Psychology: Applied*, *19*(2), 95–107. http://doi.org/10.1037/a0032696

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, *6*(1), 3–5. http://doi.org/10.1177/1745691610393980

Crump, M. J. C., McDonnell, J. V, & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS One*, *8*(3), e57410. http://doi.org/10.1371/journal.pone.0057410

Curtis, D. (2014). The man who records all his sneezes. Retrieved from http://www.bbc.co.uk/news/magazine-29525501

Dandurand, F., Shultz, T. R., & Onishi, K. H. (2008). Comparing online and lab methods in a problem-solving experiment. *Behavior Research Methods*, *40*(2), 428–434. http://doi.org/10.3758/BRM.40.2.428

Dunlop, M., Komninos, A., & Durga, N. (2014). Towards High Quality Text Entry on Smartwatches. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems* (pp. 2365–2370). http://doi.org/10.1145/2559206.2581319

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* (pp. 2669–2678). ACM Press. http://doi.org/10.1145/2207676.2208659

Evans, C., Abrams, E., & Reitsma, R. (2005). The Neighborhood Nestwatch Program: Participant Outcomes of a Citizen-Science Ecological Research Project. *Conservation Biology*, *19*, 589–594. http://doi.org/10.1111/j.1523-1739.2005.00s01.x

Fold It. (2015). Fold It. Retrieved from http://fold.it/

Fritz, T., Huang, E. M., Murphy, G. C., & Zimmermann, T. (2014). Persuasive technology in the real world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 487–496). http://doi.org/10.1145/2556288.2557383

Galaxy Zoo. (2015). Galaxy Zoo. Retrieved September 18, 2013, from http://www.galaxyzoo.org/

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. *Psychonomic Bulletin & Review*, *19*(5), 847–57. http://doi.org/10.3758/s13423-012-0296-9

Google Charts. (2015). Retrieved from https://developers.google.com/chart/

Gould, S. J. J., Cox, A. L., Brumby, D. P., & Wiseman, S. (2015). Home is Where the Lab is: A Comparison of Online and Lab Data From a Time-sensitive Study of Interruption. *Human Computation*, 45–67. http://doi.org/10.15346/hc.v2i1.4

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(28), 11116–20. http://doi.org/10.1073/pnas.1200196109

Heer, J., & Bostock, M. (2010). Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 203–212). http://doi.org/10.1145/1753326.1753357

Jennett, C., Eveleigh, A., Mathieu, K., Ajani, Z., & Cox, A. L. (2013). Creativity in citizen science: All for one and one for all. In *ACM WebSci 2013, "Creativity and Attention in the Age of the Web" Workshop*.

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 Errordiary Volunteers. *Human Computation*, *1*(2), 199–218. http://doi.org/10.15346/hc.v1i2.10

Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., … Baker, D. (2011). Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural & Molecular Biology*, *18*(10), 1175–7. http://doi.org/10.1038/nsmb.2119

Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 453–456). Florence, Italy: ACM. http://doi.org/10.1145/1357054.1357127

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*. ACM. http://doi.org/10.1145/2470654.2470684

Lab In The Wild. (2015). Retrieved from www.labinthewild.org

Li, I., Dey, A., & Forlizzi, J. (2010). A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 557–566). ACM Press. http://doi.org/10.1145/1753326.1753409

MacKenzie, I. S., & Soukoreff, R. W. (2003). Phrase sets for evaluating text entry techniques. In *CHI '03 extended abstracts on Human factors in computer systems - CHI '03* (pp. 754–755). ACM Press. http://doi.org/10.1145/765968.765971

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23. http://doi.org/10.3758/s13428-011-0124-6

Mason, W., & Watts, D. (2010). Financial incentives and the performance of crowds. *ACM SIGKDD Explorations Newsletter*, *11*(2), 100–108. http://doi.org/10.1145/1809400.1809422

Munson, S. (2012). Mindfulness, Reflection, and Persuasion in Personal Informatics. In *CHI 2012 Personal Informatics in Practice: Improving Quality of Life Through Data Workshop*. Retrieved from http://personalinformatics.org/docs/chi2012/munson.pdf

Old Weather. (2015). Old Weather.

Organisciak, P., & Twidale, M. (2015). Design Facets of Crowdsourcing. In *Proceedings of the 2015 iConference*.

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Cardamone, C., Murray, P., … Vandenberg, J. (2013). Galaxy zoo: Motivations of citizen scientists. *Astronomy Education Review*, *12*(1), 1–41.

http://doi.org/10.3847/AER2011021

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., … Vandenberg, J. (2010). Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. *Astronomy Education Review*, *9*. http://doi.org/10.3847/aer2009036

Ramsey, S. R., Thompson, K. L., Mckenzie, M., & Rosenbaum, A. (2016). Psychological research in the internet age: The quality of web-based data. *Computers in Human Behavior*, *58*, 354–360. http://doi.org/10.1016/j.chb.2015.12.049

Reinecke, K., Arbor, A., & 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* (pp. 1364–1378). http://doi.org/10.1145/2675133.2675246

Rogers, Y. (2011). Interaction design gone wild: striving for wild theory. *Interactions*, *18*(4), 58–62. http://doi.org/10.1145/1978822.1978834

Rotman, D., Preece, J., Hammock, J., Procita, K., Hanse, D., Parr, C., … Jacobs, D. (2012). Dynamic Changes in Motivation in Collaborative Citizen- Science Projects. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 1–10). http://doi.org/10.1145/2145204.2145238

Salthouse, T. A. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological Bulletin*, *99*(3), 303–319. http://doi.org/10.1037/0033-2909.99.3.303

Salthouse, T. A., & Saults, J. S. (1987). Multiple spans in transcription typing. *The Journal of Applied Psychology*, *72*(2), 187–96. http://doi.org/10.1037/0021-9010.72.2.187

Swan, M. (2012). Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. *Journal of Sensor and Actuator Networks*, *1*(3), 217–253. http://doi.org/10.3390/jsan1030217

Swan, M. (2013). The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. *Big Data*, *1*(2), 85–99. http://doi.org/10.1089/big.2012.0002

Test My Brain. (2015). Retrieved from www.testmybrain.org

The Royal Society for the Protection of Birds. (2015). The RSPB: Big Garden Birdwatch. Retrieved January 1, 2015, from https://www.rspb.org.uk/discoverandenjoynature/discoverandlearn/birdwatch/index.aspx

Von Ahn, L., & Dabbish, L. (2008). Designing Games With A Purpose. *Communications of the ACM*, (8), 58–67.