Situated Crowdsourcing Using a Market Model

Simo Hosio[1], Jorge Goncalves[1], Vili Lehdonvirta[2], Denzil Ferreira[1], Vassilis Kostakos[1]

[1] Department of Computer Science and Engineering
University of Oulu
Erkki Koiso-Kanttilan Katu 3, 90570 Oulu, Finland
firstname.lastname@ee.oulu.fi

[2] Oxford Internet Institute
University of Oxford
1 St Giles Oxford OX1 3JS United Kingdom
vili.lehdonvirta@oii.ox.ac.uk

ABSTRACT

Research is increasingly highlighting the potential for situated crowdsourcing to overcome some crucial limitations of online crowdsourcing. However, it remains unclear whether a situated crowdsourcing *market* can be sustained, and whether worker supply responds to price-setting in such a market. Our work is the first to systematically investigate workers' behaviour and response to economic incentives in a situated crowdsourcing market. We show that the market-based model is a sustainable approach to recruiting workers and obtaining situated crowdsourcing contributions. We also show that the price mechanism is a very effective tool for adjusting the supply of labour in a situated crowdsourcing market. Our work advances the body of work investigating situated crowdsourcing.

Author Keywords

Crowdsourcing; virtual currency; market; situated technologies.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

An important limitation of existing crowdsourcing markets is the lack of domain-specific expertise in the workforce [11], and the difficulty in recruiting workers from particular geographic areas for tasks that require contextualised knowledge [8]. For instance, while the creation of newspaper articles [1] and the translation of documents [34] are appealing crowdsourcing applications, they require workers within a relevant context. Mobile-based crowdsourcing systems can address this limitation, but they introduce certain constraints: costs for workers (data plan), the installation of third-party software, and in some cases consent to being tracked [20].

Situated crowdsourcing technologies have emerged in an attempt to overcome these limitations of online and mobile crowdsourcing markets [7,11]. Goncalves et al.'s situated touchscreen prototype leveraged altruism and intrinsic motivation to recruit volunteers for medical image analysis [7], and Heimerl et al. used a touch-screen vending machine to attract workers and reward them with snacks [11]. These bespoke examples highlighted the potential benefits of crowdsourcing using situated technologies, but they fail to address an important question of scaling-up into a market: *can a situated crowdsourcing market be sustained*?

A second challenge for situated crowdsourcing is managing the influx of workers' contributions, and specifically the need for contributions at a particular geographic location or a time of day [9]. How can those be encouraged, given that altruism and intrinsic motivation are challenging to manipulate precisely [7]? Economic theory suggests that a market-like platform can use price-setting to adjust the supply of different types of labour [12]. Along these lines, Heimerl et al. used a market-based approach to situated crowdsourcing, but did not examine the feasibility of using the price mechanism for control [11]. Furthermore, their economic model was inconsistent between sessions, likely limiting its power to shape user behaviour. In addition, it has been noted that intrinsic, usability, and social factors may confound the effect of price-setting [18]. Thus, a second important question is: *can the price mechanism be used to control the supply of labour in situated crowdsourcing?*

We addressed both questions – regarding sustainability and price-setting – by building a situated crowdsourcing marketplace (Bazaar) and studying it experimentally over several weeks. Bazaar uses a grid of touch-screen kiosks scattered across campus, and rewards workers with a virtual currency (HexaCoins) for completing tasks. Workers can exchange HexaCoins for money, for other goods (e.g., movie tickets), or with other workers.

Our work is the first to systematically investigate workers' behaviour and response to economic incentives in a situated crowdsourcing market. Our findings indicate that the market-based model is a sustainable approach to recruiting workers and obtaining crowdsourcing contributions. We also show that the price mechanism is a very effective tool for adjusting the supply of labour from different locations, different times of the day, and different task categories. Finally, we found that the system's virtual economy encouraged some emergent user-to-user coordination, further boosting the labour supply.

RELATED WORK

Situated Crowdsourcing

Crowdsourcing with ubiquitous technologies is increasingly gaining researchers' attention [19,32], especially on mobile phones. This has allowed researchers to push tasks to workers anywhere and anytime. Most mobile platforms

UIST '14, October 05 - 08 2014, Honolulu, HI, USA

ACM 978-1-4503-3069-5/14/10…$15.00.
http://dx.doi.org/10.1145/2642918.2647362

have been deployed in developing countries targeting low-income workers and providing them with simple tasks (e.g., [10]). Recent advances in mobile technologies have also allowed for more intricate and creative tasks. For instance, location-based distribution of crowdsourcing tasks has allowed workers to perform highly context specific tasks for others. Some examples of this include giving location-aware recommendations for restaurants [1], providing instant weather reports [1] or authoring news articles by requesting photographs or videos of certain events from workers [31].

An active community has grown around the topic of crowdsourcing measurements and sensing [19]. This participatory sensing movement is part of the larger concept of "Citizen Science" [23] that relies on mobilizing large parts of the population to contribute to scientific challenges via crowdsourcing. Often this involves the use of smartphones for collecting data [5] or even donating computational resources while one's phone is idle [3].

Despite the appeal of mobile phones, using them for crowdsourcing requires workers' implicit deployment, configuration and use of the device. For example, in SMS-based crowdsourcing, participants need to explicitly sign up for the service, at the cost of a text message exchange. This makes worker recruitment challenging, as a number of steps are necessary before a worker can actually contribute using their device. An alternative approach is to embed input mechanisms (e.g., public displays, tablets) into a physical space and leverage users' serendipitous availability [22]. This means that, contrary to mobile crowdsourcing, situated crowdsourcing through embedded interfaces does not require any deployment effort from workers [7].

In such a deployment, Heimerl et al. reported Umati [11], which used a vending machine with a touch display for locally relevant tasks, albeit with certain limitations. For example, it was available at a single location only, and it lacked diverse tasks to keep users engaged for long. Goncalves et al.'s public display crowdsourcing deployment also suffered from the lack of diverse tasks [7]. These findings suggest that task diversity is key to sustaining a situated crowdsourcing market.

Digital Markets, Currencies, and Economies

It has become increasingly common to build digital markets, currencies, and entire "virtual economies" into multi-user computational systems, in order to coordinate and direct user attention and effort in desired ways [18]. One potential application area for this approach is crowdsourcing. Market as an abstract concept refers to the interaction of supply and demand for a particular good or service resulting in exchange [18]. A market can be implemented in a computational system by creating a marketplace. Crowdsourcing marketplaces, also known as online labour markets, are a prominent example [e.g. 10,11,12,15,21]. They are used as an alternative to other mechanisms (such as appeals to altruism [7]) to motivate user contributions and to allocate tasks to different users.

Theoretically, markets have certain properties that make them attractive for this purpose. First, markets encourage users to specialize in tasks where they have a comparative advantage and to develop their skills further. Second, the types and quantities of labour supplied through a market can be controlled and adjusted through price-setting. In comparison, altruism-based contributions are hard to control and adjust, as they depend on the person and the type of task at hand (e.g. editing a favourite article on Wikipedia vs. doing uninteresting microtasks). Third, markets are sustainable over time, as long as valuable rewards can be provided. In comparison, the sustainability of contributions based on altruism, social status and similar motivations is more idiosyncratic and unpredictable. Therefore, adopting a market model for general-purpose situated crowdsourcing can enhance the sustainability of such a platform.

Suppliers in a market must be naturally rewarded with something that is of value to them. For example, Heimerl et al. used snacks to reward workers in their vending machine-based situated task marketplace [11]. However, suppliers are not usually paid directly in goods; instead, they are paid with money, which can later be exchanged for goods. Economic literature suggests that money is used as an intermediary for three basic reasons [18]. First, money is a convenient medium of exchange, as it allows the holder to choose their most preferred set of final goods out of all the possible combinations. Second, money is a convenient way of storing value over time, until the most preferred moment of consumption or until sufficient value is accumulated. Third, money provides a convenient unit for measuring value accumulation in numerical terms. Any good or record used for these three purposes can be called money; the term "currency" emphasises the first purpose [18].

Heimerl et al.'s system implements a virtual currency that can be exchanged for snacks. However, it does not allow user-to-user credit transfers or the freedom that utilizing a national currency yields [11]. Many other task marketplaces reward workers in national currency [e.g. 12], which facilitates all three uses, and also has the advantage that the platform owner does not need to provide means for exchanging the currency to goods, as national currency can be exchanged elsewhere.

However, virtual currencies have one distinct advantage over national currencies in experimental systems: the cultural, regulatory, and security-related expectations relating to them are far more open, allowing for bold experimentation [18]. Virtual currencies can also be perceived as more playful [33], though this is not necessarily an advantage in serious task market. Thus, using a virtual currency affords all three traditional functions of money and to maximize the market's effectiveness in experimental settings, as long as the currency is redeemable for valuable goods.

Finally, some systems that feature a virtual currency also make it possible for users to transfer the currency between each other [18,33]. This has two potential advantages. First, it allows users to collaborate by e.g. pooling their efforts to reach high-value rewards, potentially creating social effects that increase the labour supply. Second, it allows users to potentially start using the currency to mediate transactions completely unrelated to the original platform, in the way national currency is used. For example, users of the Tencent QQ instant messaging platform began to use its "Q Coin" currency to trade everything from compact discs to online

game items [18]. The advantage to the issuer is that this greatly increases demand for the currency, and in the case of a task market, the desirability of the tasks through which it is earned.

In summary, economic theory and practical experience highlight two important aspects for situated crowdsourcing: a market-driven model (rather than intrinsic motivation) can enhance its sustainability, and using a virtual currency (rather than directly goods) can provide enhancements to the market's operation, as long as it can be redeemed for goods. Taking these into account, we designed a market for situated crowdsourcing that uses a virtual currency, and has a redeeming mechanism to convert currency into goods. We call our market "Bazaar", and the virtual currency "HexaCoins".

SYSTEM DESCRIPTION

Bazaar is a grid of crowdsourcing kiosks (Figure 1) coordinated by an online server. Each kiosk contains an Android tablet with a 10.1" touch-screen, a charger, and uses WiFi to connect to the server. Our client software for the tablets was set to "kiosk mode" [29] to ensure that it was always visible on screen, it recovered from crashes, and unwanted OS functionality (notification bars, etc.) was disabled. The physical buttons of the tablet were physically obscured by the kiosk's enclosure.

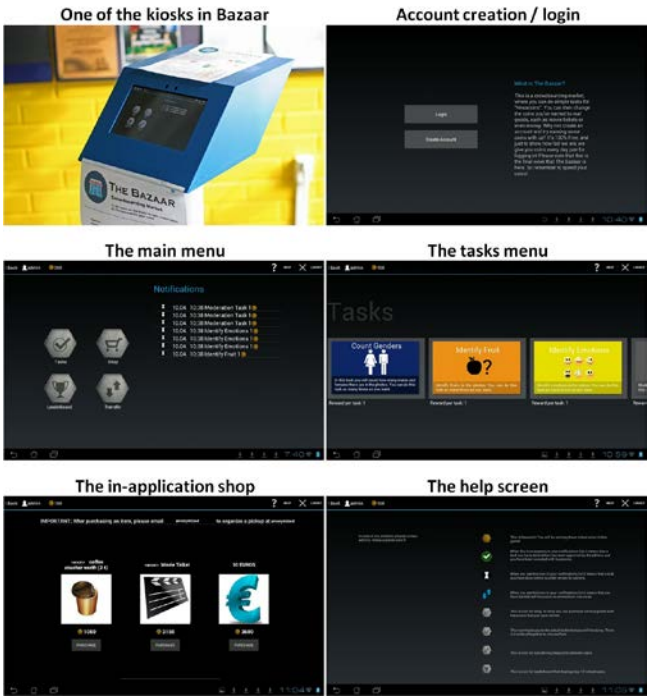

Figure 1. A close up photo of a crowdsourcing kiosk in Bazaar and screenshots of the application screens.

Kiosks

The welcome screen of the kiosks contained a brief description of the system, and prompted users to create an account or login. Registration required just a username and password because a lengthy process can reduce participation [4]. Upon login, users had to acknowledge a pop-up message that administrators can dynamically modify. Subsequently, users were shown the main screen:

- At all times the upper bar of the application displayed the HexaCoin balance of the user, and a link to the instructions window.
- Tasks allowed users to access the available crowdsourcing tasks and complete them.
- Leaderboard showed the nickname of users with most earnings to date.
- Shop allowed users to use HexaCoins to buy items.
- Transfer allowed users to send their HexaCoins to other users.
- Notifications showed confirmation that HexaCoins were credited to the account, or they were still pending moderation approval.

Participants could get money or goods in exchange for their HexaCoins by visiting the Shop page. It allowed them to purchase the desired items (cinema tickets, physical badges, tablet, money) using their HexaCoins. They subsequently had to email us to schedule the pickup from a physical shop we had setup in our laboratory premises. When visiting our lab, we asked them to verify their identity by logging into Bazaar on a dedicated computer. In addition, we used their visit as an opportunity to conduct interviews.

Tasks

We explicitly decided to provide a variety of crowdsourcing tasks to cater to the varying interests and skills of workers [15] and to keep them engaged longitudinally. Therefore we selected a diverse set of tasks with different stimuli and purposes. The workers always had the option to skip a particular task in case they did not know the answer. When a worker completed a task, the next one would automatically show up, chosen randomly from that task's repository. There were 6 different types of tasks available in Bazaar (Table 1 & Figure 2):

- **Data Categorisation**: Categorisation and labelling of photographs is a frequently offered crowdsourcing task due to its computational complexity. We provided one task where workers had to count the number of males and females in a photograph, and another where they had to type the name of the fruit shown in a photograph.
- **Sentiment Analysis**: Humans can identify at least six different emotional states (anger, happiness, sadness, fear, surprise and disgust) quite reliably [6], but the computational complexity makes this type of tasks frequent candidates for crowdsourcing. For this task we showed workers a looping 3-second video of a person's face, and asked them to identify the emotional state of the individual using six response buttons labelled with the above emotional states.
- **Content Creation**: Crowdsourcing content creation has been explored previously in the context of creating news articles [1], translating documents [34], and providing recommendations [31]. This type of task can greatly benefit from workers' local knowledge, which is key to in-situ crowdsourcing [8]. For this task, workers had to type a textual description of their surroundings. A worker could complete this task only once per Bazaar kiosk.
- **Content Moderation**: Quality control can be an overwhelming challenge for crowdsourcing [15], and one approach is to involve workers themselves to

"approve" each other's work [17]. For this task, workers had to review other workers' tasks and label them as "good" or "bad". This pool of tasks grew in real-time as workers completed tasks across all Bazaar kiosks.

- **Survey**: The survey was a one-off task that each worker could complete only once, and only after they had completed 30 other tasks. It contained a set of open-ended questions regarding how they found out about Bazaar, their motivations behind using it, any suggestions of improvements and a standardised System Usability Scale (SUS).

Task Category	Unique Tasks Available	Type	Stimulus	Worker Input	Reward (HexaCoins)
Counting Genders	373	Data categorisation (counting)	Static (images)	Text (numbers)	10
Identifying Fruits	370	Data categorisation (identification)	Static (images)	Text (short)	10
Identifying emotions	1350	Sentiment analysis	Dynamic (videos)	Multi-choice buttons (6)	5
Describing Location	4	Content Creation	Text	Text (long)	150
Moderation	Same as number of tasks approved	Content Moderation	Static, dynamic, text	Multi-choice buttons (2)	5
Survey	1	Survey	Text	Text and radio buttons	500

Table 1. Summary of number of unique tasks, types, stimuli, worker input and initial reward.

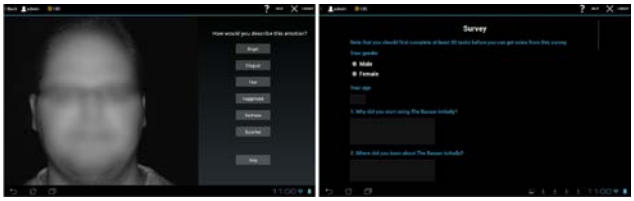

Figure 2. Examples of the tasks: identifying emotions and the in-application survey.

We chose to include both "typical" online crowdsourcing tasks and "situated" tasks. The "situated" tasks were the tasks in 3 categories: "Content Creation", "Survey", and "Moderation" (i.e. moderating the other situated tasks). Our decision aimed to enable the comparison of our results to those of online crowdsourcing studies, and at the same time to analyse how Bazaar performs with situated tasks. Platforms like Bazaar should not be built solely for situated tasks, but be regarded as complementary to already existing solutions. Thus, it is crucial to experiment with a rich variety of different tasks.

Rewards

Our rewards are HexaCoins, goods, and cash. When completing tasks, users receive HexaCoins. All completed tasks were subject to a 2-stage moderation. Administrators first rejected obviously flawed responses, and subsequently the crowd moderated the remaining responses. We decided that moderation and rewarding should be completed in chunks because per-task payment encourages gaming behaviour [15] and can reduce quality [16].

The HexaCoin value of different types of tasks varied, but we aimed to reward users with a baseline of approximately 3600 HexaCoins per hour of work. Furthermore, we rewarded users with 100 HexaCoins on the first login of each day, to motivate them to return daily and perform more tasks. Users could ultimately exchange HexaCoins for goods, using a rough exchange rate of 360 HexaCoins per 1€ (making the theoretical hourly wage 10€):

- Badges (i.e., insignia) that are popular decorations in students' ceremonial overalls (720 and 1080 HexaCoins)
- University cafeteria coffee vouchers of nominal value 3€ (1080 HexaCoins)
- Movie tickets to the local cinema (3150 HexaCoins)
- One 16Gb Android tablet (72000 HexaCoins)
- Money in the form of 10€ and 25€ packs (3600 and 9000 HexaCoins, respectively).

We avoided direct conversions between HexaCoins and money for simplicity and to encourage users to complete more tasks instead of rapidly cashing in.

Data Logging

Bazaar logged centrally all interactions on all kiosks: logins, logouts, starting and ending of performing tasks (time spent), answers for each task, HexaCoin transfers, responses to the survey. All users who bought goods from Bazaar were interviewed when they picked up their purchases from the shop set up at our laboratory. The interview consisted of open-ended questions about their motives to use Bazaar, if they collaborated with other users, why did they choose the particular rewards they purchased, would they continue to use the system if deployed for longer, how did they feel about having to perform tasks in such public spaces, and if they had prior experience with crowdsourcing in general.

STUDY

We designed a real-user experiment to investigate if Bazaar enabled the creation of a crowdsourcing market in the economic sense. Our experiment systematically manipulated the amounts of HexaCoins paid in reward at specific locations (week 1, 2); for specific tasks (week 2, 3); and during specific times of day (week 3), in an attempt to investigate if the labour supply responded accordingly.

We deployed four Bazaar kiosks across the University of Oulu (in Oulu, Finland) campus for 3 weeks. We did not actively promote Bazaar except by attaching an A3-sized poster on each of the kiosks. We specifically avoided the

use of email lists, Facebook and Twitter, to minimise participation bias.

The locations where we deployed Bazaar were physically afar and at different faculty buildings (Figure 3):

- Location 1: a cafeteria visited mostly by technical students
- Location 2: next to the biggest on-campus restaurant
- Location 3: a lobby area with benches surrounding the kiosk
- Location 4: next to the main entrance of the main library

Figure 3. Bazaar deployment locations. Top row: cafeteria and next to the main restaurant. Bottom row: a lobby with benches and next to a library entrance.

During the study 194 accounts were created, 1067 logins, 75229 tasks completed (62602 approved) in 310114 seconds (86.1 hours) of crowdsourcing effort, and 832548 HexaCoins generated (Figure 5). The most popular task category was moderation (N=23986), followed by counting genders (N=14011), identifying emotions (N=13624) and identifying fruits (N=10765). On the other hand, the location description task was completed 138 times and the survey 78 times. A total of 25 transfers were registered (to 10 unique users) worth 14600 HexaCoins in total.

Of the 194 accounts created, 97 (50%) were returning users. Furthermore, 87% of the HexaCoins redeemed, while the remaining 13% (110014 HexaCoins) were simply unused. We also analysed how the HexaCoins generated within Bazaar were distributed amongst workers. Figure 4 depicts the relationship between HexaCoins and time spent working. It shows that workers enjoyed increasing returns on time spent working, suggesting that they developed their skills over time, just as in a conventional labour market.

"Happy Place" Manipulation of Rewards

Our experiment sought to investigate whether workers can be motivated to change their location, and complete tasks at particular kiosks on request. To test this manipulation we introduced a reward multiplier, applied to one of the kiosks at a time. For the duration of a whole day, a single kiosk (i.e. the "Happy Place") yielded twice (2x) the HexaCoins for each task completed, while all other kiosks operated as usual. We applied this manipulation on four sequential days (Mon-Thu), each day with a different Happy Place.

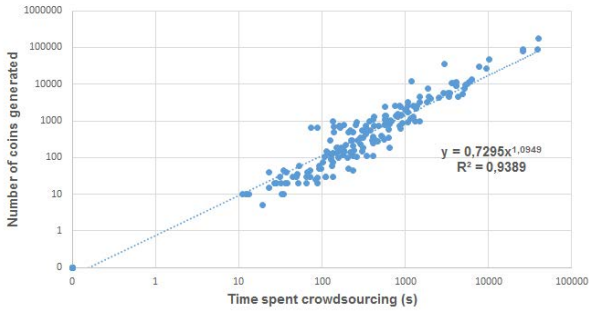

Figure 4. Scatterplot of how many HexaCoins (y-axis) and effort (x-axis) are attributed to each worker (data points). Both axes are in logarithmic scale.

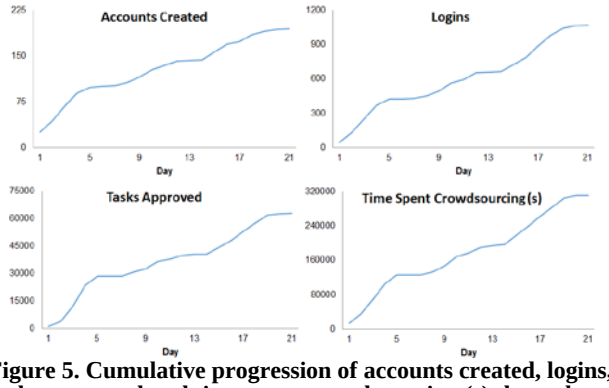

Figure 5. Cumulative progression of accounts created, logins, tasks approved and time spent crowdsourcing (s) throughout the deployment.

Figure 6 shows the breakdown of time spent by workers at each location when the Happy Place effect was active. For Monday-Thursday we observe that workers spent more time at the "Happy Place" than other locations by a factor of 4.02 on average. Friday did not have a happy place and is shown as a baseline assessment.

In the subsequent week we reduced the Happy Place multiplier from 2x to 1.5x, expecting to observe a reduced effect. In Figure 7 we show that, as expected, "Happy Places" were more popular than other locations by a factor of 3.77 on average. Thursday did not have a happy place and is shown as a baseline assessment.

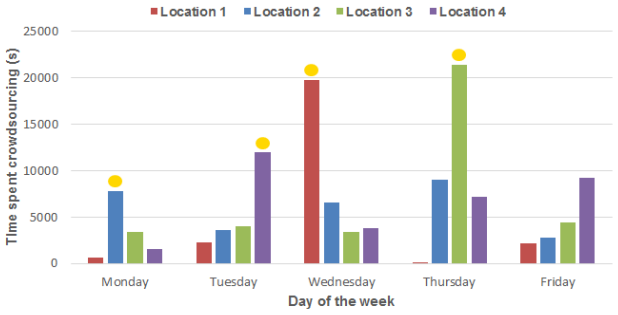

Figure 6. Crowdsourcing effort (seconds) at each location. "Happy Place" multiplier (2x) was used in locations with the yellow marker.

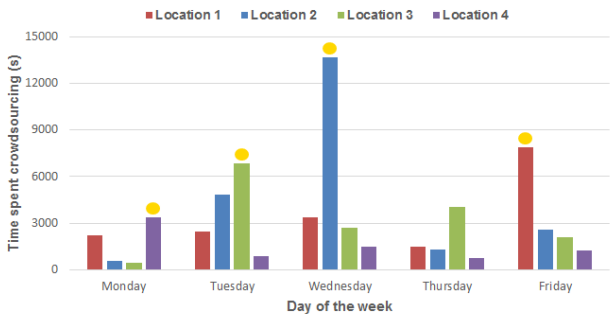

Figure 7. Crowdsourcing effort (seconds) at each location. "Happy Place" multiplier (1.5x) was used in locations with the yellow marker.

Manipulation of Task Rewards

During the first week of deployment the workers completed the tasks shown in Figure 8, indicating a strong preference for tasks in the Moderation category. We sought to investigate whether we can manipulate the popularity of tasks by manipulating their reward. Our manipulations had the expected result: an increase in HexaCoin reward yielded an increase in popularity, and vice versa.

During week 2 we modified the rewards per category (Table 1) as follows:

- The reward for tasks in the Moderation category was reduced from 5 to 2 HexaCoins. This yielded a 7-fold decrease in popularity during week 2.
- The reward for tasks in the Identifying Emotions category increased from 5 to 10 HexaCoins. This yielded a 3-fold increase in popularity during week 2.

During week 3 we made further manipulations to the rewards per category as follows:

- The reward for tasks in the Identifying Fruits category decreased from 10 to 5 HexaCoins. This yielded a 4-fold decrease in popularity during week 3.
- The reward for tasks in the Counting Genders category increased from 10 to 15 HexaCoins. This yielded a 10-fold increase in popularity during week 3.

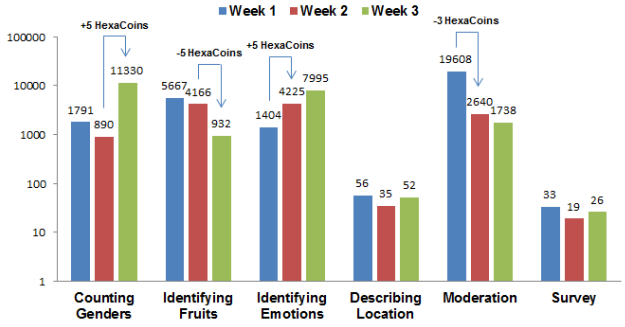

Figure 8. Number of tasks completed per category during each week of deployment (y-axis in logarithmic scale). The arrows indicate where a change in reward was done.

"Happy Hour" Manipulation of Rewards

During the first two weeks of deployment we observed that workers mostly completed tasks between 10am and 7pm (Figure 9). We investigated whether workers could be motivated to change the time when they perform tasks in Bazaar. To test this we introduced a "Happy Hour" reward multiplier (2x) applied to all kiosks simultaneously at certains times: 9-10am, and 8-9pm. These happy hours were chosen because they were relatively unpopular during the first two weeks. As expected, "Happy Hours" were significantly more popular during week 3 (Figure 9). Specifically, there was a 2-fold popularity increase between 9-10am, and an 8-fold popularity increase between 8-9pm.

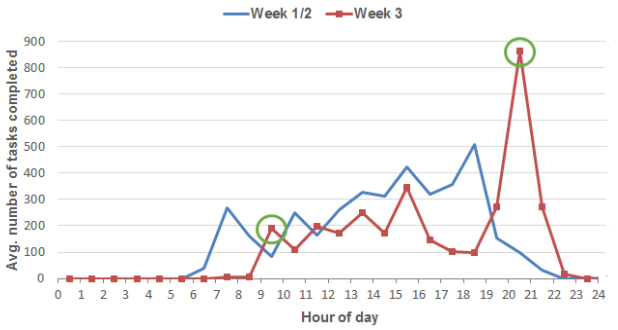

Figure 9. Average number of tasks complete per hour during the first 2 weeks (blue) and the third week (red). "Happy hours" are indicated with a circle.

Surveys and Interviews

In total 78 users completed the Bazaar survey task (51 male, 27 female). The average age was 23.8 (SD=4.1). The noted three key reasons for why they started using Bazaar: to get the rewards illustrated in the posters (N=22); they were recommended by a friend (N=14); or out of curiosity (N=41). One respondent did not answer. When asked where they learnt about it we identified two main responses: either the respondents indicated that they just stumbled upon the kiosks at the campus (N=55), or they were informed by their friends (N=23). The improvement suggestions we collected are summarised in the discussion.

Analysis of the SUS revealed a score of 81.3 (SD=10.8) on a scale from 0 to 100. The positive statement with the lowest value for positive was if users would like to use the system frequently (M=3.6, SD=1.2). Other values showed that users did not consider the system to be complex (M=1.8, SD=0.7), found that it was easy to use (M=4.4, SD=0.7), can quickly be learned (M=4.4, SD=0.7), and requires no technical support (M=1.2, SD=0.5). To provide a fairer grading assignment, we used percentiles like those calculated in [26] using a curved grading scale. This means that the SUS score for our system obtained an A grade (above 80.3%).

Finally, 45 workers (26 male, 19 female) of the 194 who created an account in Bazaar purchased prizes and were interviewed during their pick-up of the items. The average age was 23.9 (SD=3.8). The key findings from the interviews are used to support our discussion.

DISCUSSION

Situated crowdsourcing differs from online crowdsourcing substantially. Therefore, we do not consider it an alternative but a complementary means of enabling crowd work, which addresses certain limitations of online crowdsourcing [15].

Our study is the first in-depth investigation of how workers behave in a situated crowdsourcing market. Previous studies on situated crowdsourcing lacked task diversity,

reward flexibility, and crucially the ability to track the behaviour of individual workers [7,11]. Our study addresses these limitations, and provides a holistic assessment of worker behaviour in a situated crowdsourcing market.

There are three criteria we use to judge the success of our study from a crowdsourcing perspective: attractiveness, speed, and quality. In addition, we assess the system's sustainability, the extent to which we were able to control the labour supply using the price mechanism, and the extent to which the virtual economy encouraged the emergence of advantageous user-to-user dynamics in the system.

We first highlight an important finding from our interviews: the majority of workers in Bazaar were completely new to crowdsourcing, and in interviews admitted to have never used any of the popular crowdsourcing markets such as Amazon's Mechanical Turk and Crowdflower. This strongly indicates the potential of situated crowdsourcing to reach untapped populations of workers.

Second, we point out that the number of tasks completed in Bazaar remained at a constantly high level: we approved 62602 tasks, or almost 3000 tasks per day. Task completion rate in situated crowdsourcing has been found generally greater than the completion rate of tasks on MTurk. For instance, Rogstadius et al. purchased labour from MTurk for a simple image analysis task, resulting in approximately 40 daily completed tasks [26]. Further, the performance of *Umati* was compared to MTurk, finding the situated approach with only a single deployed interface as capable of producing over 3x more daily labour at the rate of over 1000 tasks per day [11]. While MTurk studies with high task throughput exist [25], we feel that Bazaar with only four deployed kiosks achieved a workforce throughput that is at least comparable to MTurk and previous situated crowdsourcing studies.

Third, the quality of completed tasks was also high. We relied on crowd-based quality assurance, which has been reported as a viable mechanism [8,17]. In Bazaar, the moderation task resulted in 94% of positive votes, indicating high quality of the performed tasks. We also note that only a handful of workers abused the system by completing tasks in a negligent manner, which is expected when reward is per task rather than per hour [15]. However, our moderation and rejection of bad quality work did substantially curb Bazaar's abuse, and in fact we did notice abusive workers eventually produced high quality work.

Bootstrapping and Sustaining the Workforce

Worker acquisition is one of the key challenges for crowdsourcing [11]. Previous studies on situated crowdsourcing have highlighted its promising ability to rapidly attract capable workers [7,8]. Similarly in our study, Bazaar managed to quickly attract a large number of unique users (194) to create accounts and contribute work. The kiosks in four different appealed to a heterogeneous mixture or both males and females, students and staff, and from different faculties.

We note that the frequency of logins, time spent in performing tasks, and the number of tasks completed remained high until the end of the study (Figures 5-8). The surveys we deployed in Bazaar revealed the major role of curiosity in people's motives to start using the kiosks, and interview comments such as "*I use all new 'gizmos' around me out of pure interest, and then I realized that I can actually earn something using this!*" and "*I just happened to walk by during lectures and had free time to start exploring what it was*" do support this. This agrees with earlier research stating that public situated technologies are often initially used out of curiosity rather than with a clear motive in mind [22] and that they are thus capable of reaching broad audiences [13].

We also note that we did not actively promote Bazaar, and that many of the interviews and surveys noted that word-of-mouth was key in attracting new workers. In fact, the promotion of Bazaar was sustained by workers themselves. Many of the survey respondents indicated being recommended by their friends to start using it, and interviews further revealed that workers used Facebook, email and SMSs to inform their friends about Bazaar. This is crucial because public deployments often find it challenging to entice commitment from new users unless it clearly benefits them [4]. In our case, the value proposition to potential users was appealing and made credible by recommendations from friends. The promise of "real value" was constantly mentioned in the interviews, and especially when workers claimed rewards and informed (often bragged) about it to their friends, the appeal of Bazaar seemed to grow, leading into sustained use. Interview comments such as "I saw a friend claiming 20€ from it and wanted to try as well" or "One friend came with an overall patch and said that anyone can earn these things" illustrate how word-of-mouth promotion by the workers themselves was generating awareness about Bazaar.

An important factor in Bazaar's attractiveness was the diversity of offered tasks. Previous research has shown that the lack of task diversity can be detrimental [10,11]. Many mobile and ubiquitous crowdsourcing systems have been research-driven and have not involved stakeholders who are actually willing to pay for tasks to be completed. As a result, most systems involve rather unappealing tasks. As such, research has shown that workers lose interest quite rapidly [10], even when rewards are monetary. In Bazaar we attempted to provide a rich set of tasks so that workers would always have a diverse choice, even if that meant completed a task more than once. Some workers claimed in their interview to be annoyed by this, but most did not mind.

Controlling Labour Supply through Price-setting

One of the key aims of our experiment was to test whether price-setting can control labour supply. In small tasks such as those done in crowdsourcing, it is not obvious whether workers are motivated more by rewards or by intrinsic motivations such as altruism or the novelty of the system. In the latter case, increasing the rewards could fail to increase the contributions, or even decrease them [10]. A few studies have tested this in Mechanical Turk and found that in this context, higher rewards are indeed associated with more work delivered, but not necessarily of higher quality [26]. In the context of Bazaar the results could be different, because the contributors did not actively seek earnings opportunities, but instead the system was placed so as to capture contributors with the desired situated information. Thus it was important to test whether the price

mechanism could be used to control the relative quantities of different types of contributions.

We systematically manipulated rewards to investigate if worker behaviour can be adjusted in terms of which tasks they perform, where they work, and when they work. Our results confirm that reward manipulation led to a predictable and proportional effect in terms workers preference of tasks (Figure 8), crowdsourcing location (Figures 6 and 7), and time of day when they worked (Figure 9). In economic terms, labour suppliers in the Bazaar marketplace exhibited high price elasticity, making it very feasible to control and adjust their contributions through the price mechanism.

When we asked the users whether they attempted to take advantage of the reward multipliers, 4 said no, while 40 claimed to have been motivated by them to adapt their working behaviour to earn Hexacoins faster: "*Yesterday I planned to. I had 30 minutes of free time, so I walked to a place with the multiplier [happy place], but it was occupied. I wanted to do the "survey" and "describe location" tasks somewhere with a higher multiplier, so I saved them for a later time*", "*Yes, I started randomly on whichever kiosk was close by and got lucky a few of times, starting immediately on correct one. Other times I tried to relocate, but they were occupied!*", or "*Yes, very much so. I moved to the locations with multipliers always when possible and when they were not occupied.*" Several quotes like this suggest that we could have scaled the system up by simply deploying more kiosks than just one per location. This behaviour can also be seen in our log data, but exact quantification is impossible as many users indicated that they used a kiosk with the multiplier multiple times by chance.

However, the price mechanism was not without limits. The interviews revealed that the morning "happy hour" was a somewhat misguided intervention because it coincided with lectures' schedule, or it was too early in the day for our workers. As a result, the yield of the morning "happy hour" was much less than from the evening one. This highlights the importance of understanding context when working with situated technology. In economic terms, opportunity costs created by prior commitments and contextual needs could in some cases be higher than the rewards offered by the market.

As expected, some interviewees also indicated that they were not motivated by the currency rewards as such, but by the goods that the currency could be exchanged for: "*I would never choose money. I rather have a more rewarding experience of getting something tangible...something that I will more likely remember in the future*," "*I have exactly two movies that I want to see, and this way I know exactly where I will use the reward*," or "*I love my student overalls and always want more patches for it. Money does not appeal to me.*" These statements also echo an important finding in economic sociology that all money is in fact not equal, as economic theory suggests [35]. People do mental accounting where they e.g. allocate money earned from a side job to an "entertainment budget", potentially making those earnings more desirable than the "boring money" earned from one's primary job.

Finally, at the end of the experiment 13% of the HexaCoins were not redeemed. In pre-paid value business (e.g. gift cards, calling cards), such unredeemed currency is known as "breakage". Breakage is seen as a natural phenomenon, and contributes significantly to profits. In a digital labour market such as Bazaar, breakage means free labour, but it might also indicate a problem with sustainability, especially if many workers leave their rewards unredeemed because they disappear from the system.

Emergent User-to-user Dynamics

Our assessment of workers' use of Bazaar involved a situated SUS survey. Previous research has shown that systems scoring above 80.3% (grade A) in SUS are more likely to be recommended by users to their friends [28]. Our analysis indicated that Bazaar scored 81.3%, just above the identified tipping point. This suggests that workers' use of Bazaar was enjoyable enough to promote to their friends. This is important because situated technologies are often used in groups, since the social pressure in public interaction [4] is lower when users are in groups [13].

Recent work has hypothesized that for crowdsourcing it might be beneficial to attract loners, i.e. individual users instead of groups, because the quality of their work is better than that of groups [7]. Our interviewees claimed that they used Bazaar practically always alone: it made no sense to perform tasks in groups given that only one would be rewarded. However, friends did form virtual groups, to work together towards a joint goal. For instance, the interviews revealed that a group of 4 workers decided to earn 2 movie tickets each to go see movies together. They collaborated very effectively and took advantage of happy places and hours: "*I even received SMS notifications from friends, saying where or when to go work today for maximum coins.*" 1 worker failed to notice the possibility to benefit from the multipliers.

Collaboration in the form of currency exchange, however, did not take place often. Based on our interviews the majority of the 25 recorded exchanges were motivated by curiosity, or workers quitting and transferring their HexaCoins to friends. Takayama & Lehdonvirta obtained similar results in an experimental virtual economy that allowed user-to-user transfers [30]. In contrast, user-to-user transfers have been extremely popular in some commercial virtual currency systems, and resulted in significant emergent behaviours [18]. These effects remain to be captured and exploited in a research setting.

Finally, many interviewees referred to Bazaar as a "game." This was surprising because it was not designed as such, but our interviews suggest that particularly the leaderboard influenced workers' perception: "*Yes, I guess it raised my competitive spirit, it's always nice to be #1, especially if there's not much to do to reach the position!*" and "*Yes, <another username> was right on my heels. Every morning when I came here she was ahead of me, which just motivated me to grind more. She eventually gave up, it seems (laughing).*" The leaderboard clearly motivated people to work more, as other studies on gamified crowdsourcing show [8], but at the same time it does shape people's perceptions.

Administering a Situated Crowdsourcing Market

In comparison to online crowdsourcing markets, situated markets have an additional key stakeholder: location managers [14]. These are the individuals who are involved in managing the physical location where a kiosk may be installed. While in developing countries Internet cafes are often a physical location where crowdsourcing work takes place, this aspect is amplified with situated markets. The commitment of location managers is crucial to the success of situated crowdsourcing. Therefore, the kiosks must provide value not only to workers, but also to the location managers who install them [14]. A straightforward way would be to give the managers a small percentage of the profits originating from the tasks performed at their location. This would also motivate the managers to maintain and advertise the opportunity to their audiences.

Our findings on manipulation can prove useful in a number of use cases benefiting location managers. For instance, an establishment that houses a crowdsourcing kiosk (e.g. pub or cafe) could willingly pay for the overhead in introducing reward multipliers to physically attract workers. In this sense, the platform could be used for attracting people to spaces and raising awareness of that space. Similarly, manipulating the rewards per different times of day can be used to avoid physical congestion at a particular kiosk by distributing workers' effort.

Further, the use of a virtual currency allows location managers to offer goods that their establishment actually provides (e.g. cake at a coffee house, beer at a pub) for the work performed at their premises. Utilizing a virtual currency for sustained engagement is key here, as it enables workers to work intermittently, and build towards their desired rewards. Given our findings on how money does not appeal to all workers, and the interviewees' feelings of getting "something for nothing" from Bazaar, this could benefit both the dwellers of an establishment as well as the managers.

Rewarding workers in situated crowdsourcing remains a relatively unexplored issue. On MTurk the average salary per hour is $5-7$. Therefore, in our case it is reasonable to question whether our hourly rate of roughly 10€ was too high. In Finland, where the experiment was conducted, 10€ is much lower than the average wage. The purchasing power of 10€ in our country is lower than a $7 hourly wage in cheap labour countries where workers on many online marketplaces come from. We suggest that it makes sense to adopt the price-setting in situated crowdsourcing to the contextual and cultural factors. Rewards in situated crowdsourcing cannot blindly follow online (e.g. MTurk prices) but should instead be influenced by the location of the kiosks.

We do not claim that the presented situated crowdsourcing deployment on our campus is a panacea. We argue that our crowdsourcing kiosks could be replicated at other locations that would be targeted for their potential workforce. We envision, for example, MTurk having an option to choose the types of locations where tasks are deployed.

Limitations

We acknowledge certain limitations in the presented study. We encountered run-time problems particularly with WiFi connectivity, leading to suboptimal user experience at times. This is however to be expected with any real-world deployment, and the outages usually lasted just a few minutes. The length and magnitude of the deployment, we feel, counterbalances the issue. Bazaar's current design is also not scalable, as there is only one centralized administrator view for moderating the tasks. Naturally all task providers should have their own accounts and views to the completed tasks. Similarly, we need to create facilities to input tasks runtime to the system. Finally, cultural issues were not investigated, which could affect the acceptability of situated crowdsourcing.

CONCLUSION

Bazaar is the first situated crowdsourcing platform using a market model. Over a period of three weeks, Bazaar attracted 194 workers who completed 75229 tasks in 86.1 hours of work. A clear majority of workers had no prior experience with obtaining rewards from online crowdsourcing. We demonstrate that if a platform is willing to make an effort to establish physical presence in certain locations, it will attract a populous workforce with comparable work quality to its online counterparts and a higher task uptake.

Our systematic manipulations show that Bazaar's price mechanism was capable of controlling the supply of different types of labour it produced. Specifically, we demonstrated that rewarding mechanism affected workers' preference on what kind of work to produce, where, and when to produce it. Moreover, a virtual currency was demonstrated as key enabler to sustaining workers' engagement and interest in Bazaar.

ACKNOWLEDGEMENTS

This work was supported by Microsoft Research through its PhD Scholarship Programme. Also the financial support of the Finnish Funding Agency for Technology and Innovation and Academy of Finland is gratefully acknowledged. This work is partly funded by EU grant agreement PCIG11-GA-2012-322138. We also thank Nemanja Vukota, Teemu Partanen, Miika Keisu, and Joonas Nikula for their technical contributions.

REFERENCES

1. Alt, F., Shirazi, A., Schmidt, A., Kramer, U. and Nawaz, Z. Location-based crowdsourcing: extending crowdsourcing to the real world. Proc. NordiCHI 2010, ACM (2010), 13-22.
2. Ariely, D. Predictably Irrational: The Hidden Forces That Shape Our Decisions. London: HarperCollins (2009).
3. Arslan, M.Y., Singh, I., Singh, S., Madhyastha, H.V., Sundaresan, K. and Krishnamurthy, S.V. Computing while charging: building a distributed computing infrastructure using smartphones. Proc. CoNEXT 2012, ACM (2012), 193-204.
4. Brignull, H. and Rogers, Y. Enticing people to interact with large public displays in public spaces. Proc. INTERACT, Vol. 3 (2003), 17-24.
5. Burke, J.A., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S. and Srivastava, M.B. Participatory sensing. In Workshop on World-Sensor-

Web: Mobile Device Centric Sensory Networks and Applications at Sensys (2006).

6. Ekman, P. and Friesen, W.V. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology 17, 2 (1971), 124-129.
7. Goncalves, J., Ferreira, D., Hosio, S., Liu, Y., Rogstadius, J., Kukka, H. and Kostakos, V. Crowdsourcing on the Spot: Altruistic use of Public Displays, Feasibility, Performance, and Behaviours. Proc. UbiComp 2013, ACM (2013), 753-762.
8. Goncalves, J., Hosio, S., Ferreira, D. and Kostakos, V. Game of Words: Tagging Places through Crowdsourcing on Public Displays. Proc. DIS 2014, ACM (2014), 705-714.
9. Goncalves, J., Pandab, P., Ferreira, D., Ghahramani, M., Zhao, G., Kostakos, V. Projective Testing of Diurnal Collective Emotion, Proc. Ubicomp 2014, ACM (2014).
10. Gupta, A., Thies, W., Cutrell, E. and Balakrishnan, R. mClerk: enabling mobile crowdsourcing in developing regions. Proc. CHI 2012, ACM (2012), 1843-1852.
11. Heimerl, K., Gawalt, B., Chen, K., Parikh, T. and Hartmann, B. CommunitySourcing: engaging local crowds to perform expert work via physical kiosks. Proc. CHI 2012, ACM (2012), 1539-1548.
12. Horton, J. J., and Chilton, L. The Labor Economics of Paid Crowdsourcing. Proc. ACM-EC 2010, ACM (2010), 209-218.
13. Hosio, S., Goncalves, J., Kostakos, V., Riekki, J. Exploring Civic Engagement on Public Displays. In S. Saeed (Ed.), User-Centric Technology Design for Nonprofit and Civic Engagements. Springer International Publishing (2014), 91-111.
14. Hosio, S., Goncalves, J., Kukka, H., Chamberlain, A., Malizia, A. What's in it for me: Exploring the Real-World Value Proposition of Pervasive Displays. Proc. PerDis'14, ACM (2014), 174-179.
15. Kittur, A., Nickerson, J.V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M. and Horton, J. The future of crowd work. Proc. CSCW '13, ACM (2013), 1301-1318.
16. Kochhar, S., Mazzocchi, S. and Paritosh, P. The anatomy of a large-scale human computation engine. Proc. HCOMP 2010, ACM (2010), 10-17.
17. Lampe, C., Zube, P., Lee, J., Park, C.H. and Johnston, E. Crowdsourcing civility: A natural experiment examining the effects of distributed moderation in online forums. Government Information Quarterly 31, 2 (2014), 317-326.
18. Lehdonvirta, V. and Castronova, E. Virtual Economies: Design and Analysis. Cambridge, MA: MIT Press (2014).
19. Liu, Y., Lehdonvirta, V., Alexandrova, T. & Nakajima, T. (2012) Drawing on Mobile Crowds via Social Media. Case UbiAsk: Image Based Mobile Social Search Across Languages. Proc. Multimedia Systems 18, 1 (2012), 53-56.
20. LocalMind, URL: http://www.localmind.com/, cited 04/04/2014
21. Martin, D., Hanrahan, B.V., O'Neill, J. and Gupta, N. Being a turker. Proc. CSCW '14, ACM (2014), 224-235.
22. Müller, J., Alt, F., Michelis, D. and Schmidt, A. Requirements and design space for interactive public displays. Proc. Multimedia 2014, ACM (2010), 1285-1294.
23. Paulos, E., Honicky, R.J. and Hooker, B. Citizen science: Enabling participatory urbanism. In M. Foth (ed), Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City. IGI Global (2008), 414-436.
24. Quinn, A.J. and Bederson, B.B. Human computation: a survey and taxonomy of a growing field. Proc. CHI 2011, ACM (2011), 1403-1412.
25. Rashtchian, C., Young, P., Hodosh, M. and Hockenmaier, J. Collecting image annotations using Amazon's Mechanical Turk. Proc. NAACL HLT 2010 Workshop CSLDAMT '10, ACL (2010), 139-147.
26. Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J. and Vukovic, M. An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets. Proc. ICWSM 2011, AAAI (2011), 321-328.
27. Sauro, J. A Practical Guide to the System Usability Scale (SUS): Background, Benchmarks & Best Practices. Measuring Usability LLC, Denver, (2011)
28. Sauro, J. http://www.measuringusability.com/sus.php, cited: 08/04/2014
29. SureLock | Mobile Device Lockdown. http://www.42gears.com/surelock/, cited 04/04/2014
30. Takayama, C. & Lehdonvirta, V. EcoIsland: A System For Persuading Users To Reduce CO2 Emissions. In Pervasive Persuasive Technology and Environmental Sustainability Workshop, Proc. Pervasive 2008 Workshop Proceedings, Springer (2008), 113-119.
31. Väätäjä, H., Vainio, T., Sirkkunen, E. and Salo, K. Crowdsourced news reporting: supporting news content creation with mobile phones. Proc. MobileHCI 2011, ACM (2011), 435-444.
32. Vukovic, M. and Kumara, S. Second international workshop on ubiquitous crowdsourcing: towards a platform for crowd computing. Proc. UbiComp 2011, ACM (2011), 617-618.
33. Wang, Y. and Mainwaring, S. Human-Currency Interaction: Learning from Virtual Currency Use in China. Proc. CHI 2008, ACM (2008), 25-28
34. Zaidan, O.F. and Callison-Burch, C. Crowdsourcing translation: Professional quality from non-professionals. Proc. Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, ACL (2011) 1220-1229.
35. Zelizer, V. A. The Social Meaning of Money, 2nd ed. Princeton: Princeton University Press (1994).