OULU 2015
C 535

ACTA UNIVERSITATIS OULUENSIS

Jorge Goncalves

SITUATED CROWDSOURCING: FEASIBILITY, PERFORMANCE AND BEHAVIOURS

UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU,
FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING,
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING;
INFOTECH OULU

ACTA UNIVERSITATIS OULUENSIS
C Technica 535

JORGE GONCALVES

SITUATED CROWDSOURCING: FEASIBILITY, PERFORMANCE AND BEHAVIOURS

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu for public defence in the OP auditorium (L10), Linnanmaa, on 14 August 2015, at 12 noon

UNIVERSITY OF OULU, OULU 2015

Acta Univ. Oul. C 535, 2015

Supervised by
Professor Vassilis Kostakos

Reviewed by
Assistant Professor Aniket Kittur
Associate Professor Panagiotis Ipeirotis

Opponent
Professor Albrecht Schmidt

ISBN 978-952-62-0849-7 (Paperback)
ISBN 978-952-62-0850-3 (PDF)

ISSN 0355-3213 (Printed)
ISSN 1796-2226 (Online)

Cover Design
Raimo Ahonen

JUVENES PRINT
TAMPERE 2015

Goncalves, Jorge, Situated crowdsourcing: feasibility, performance and behaviours.
University of Oulu Graduate School; University of Oulu, Faculty of Information Technology and Electrical Engineering, Department of Computer Science and Engineering; Infotech Oulu
Acta Univ. Oul. C 535, 2015
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

This thesis focuses on a systematic assessment of the *feasibility* and *performance* of situated crowdsourcing, as well as a basic understanding of the *behaviours* of its workers. While these aspects have been extensively studied for online and mobile crowdsourcing, this is not the case for situated crowdsourcing mainly due to its relative novelty. Such an assessment is crucial for the development of the crowdsourcing research agenda, so that task requesters and researchers alike can leverage, whenever appropriate, situated technologies for crowdsourcing efforts with more confidence.

The key findings of this thesis illustrate how appropriately designed crowdsourcing tasks can perform well even in a complex deployment setting: situated technologies in public spaces. In the articles presented in this thesis, we empirically demonstrate that situated crowdsourcing performance can compete with other means of collecting crowd contributions. While situated technologies have been reported in the past to suffer from credibility and misappropriation issues, one should not forego the use of these technologies for crowdsourcing purposes assuming that the tasks are not haphazardly designed. The thesis also explores the behaviours of situated crowdsourcing workers through in-situ observations, video analysis and longitudinal individual tracking.

Towards the end of the thesis, we revisit the research questions put forth in the thesis, and highlight how they were answered. We then discuss the benefits and drawbacks of situated crowdsourcing, and the differences between using non-personal and personal devices for this purpose. In both cases, the decisions made by the task requesters or researchers will ultimately depend on their goals and the task itself. We conclude the thesis by restating the thesis' research agenda, reflecting on the challenges and opportunities of situated crowdsourcing, and our future work within this area.

Keywords: behaviours, crowdsourcing, feasibility, performance, public displays, situated technologies

Goncalves, Jorge, Paikkasidonnaisen joukkoälyn tehokkuus, käyttökelpoisuus ja käyttäytymismallit.
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Tieto- ja sähkötekniikan tiedekunta, Tietotekniikan osasto; Infotech Oulu
Acta Univ. Oul. C 535, 2015
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Tämä väitöskirja keskittyy paikkasidonnaisen joukkoälyn käyttökelpoisuuden sekä tehokkuuden järjestelmälliseen arviointiin. Väitöskirja pyrkii myös ymmärtämään joukkoälyjärjestelmien työntekijöiden käyttäytymistä ohjaavia tekijöitä alustavalla tasolla. Paikkasidonnaisuus on tekijä, jota useimmiten verkossa tai mobiililaitteissa tehtävässä joukkotyötutkimuksessa ei ole mahdollista ymmärtää perusteellisesti. Paikkasidonnaisuus muodostaa kuitenkin elintärkeän osan joukkoälytutkimuksessa, ja sitä hyväksikäyttämällä sekä joukkotyön teettäjät että joukkoälytutkijat voivat soveltuvissa tilanteissa hyödyntää paikkasidonnaisia teknologioita luotettavamman joukkoälytiedon tuottamiseen.

Väitöskirjan keskeisimmät löydökset osoittavat kuinka tarkoituksenmukaisesti toteutetut joukkoälytehtävät tuottavat luotettavaa tietoa, jopa monimutkaisissa käyttöympäristöissä kuten paikkasidonnaisia teknologioita hyödyntävissä julkisissa tiloissa. Väitöskirjan artikkelit osoittavat empiirisesti paikkasidoinnaisen joukkoälyn olevan kilpailukykyinen muiden joukkoälyteknologioiden kanssa, vaikka paikkasidonnaisten teknologioiden on aiemmin osoitettu kärsivän uskottavuuden puutteesta sekä väärinkäytöstä. Tämän väitöskirjan löydökset osoittavat, että oikein suunnitellut joukkoälytehtävät sopivat hyvin käytättäväksi kyseisten teknologioiden kautta. Suorituskyvyn sekä tehokkuuden lisäksi väitöskirjassa esitellään empiirisiin havaintoihin, videoanalyysiin, sekä yksilöiden pitkäkestoiseen tutkimukseen pohjautuvia löydöksiä joukkoälytyöntekijöiden käyttäytymismalleista.

Väitöskirjan loppuosa käsittelee henkilökohtaisten ja julkisten laitteiden hyötyjä sekä haittapuolia suhteessa paikkasidonnaisten joukkoälyjärjestelmien käyttöön. Löydökset osoittavat, että kummassakin tapauksessa laitetyypin valintaa ohjaavat joukkoälytyön teettäjien tai tutkijoiden asettamat tavoitteet, sekä kyseessä olevat joukkoälytehtävät. Väitöskirjan päätteeksi palataan asetettuihin tutkimuskysymyksiin sekä vastaaviin löydöksiin, ja pohditaan paikkasidonnaisen joukkoälyn tuomia haasteita, mahdollisuuksia sekä tulevaisuuden visioita.

Asiasanat: joukkoäly, julkiset näytöt, käyttäytymismallit, käyttökelpoisuus, paikkasidonnaiset teknologiat, tehokkuus

Acknowledgements

Foremost, I would like to express my gratitude towards my supervisor, Professor Vassilis Kostakos. Our talks and discussions were always great sources of inspiration, and the work presented in this thesis greatly benefited from them. Without his guidance and belief in my capabilities, this thesis would have not been possible. I am also grateful to Professor Timo Ojala for serving as my follow-up group chair and providing valuable feedback over the years, Dr. Mika Rautiainen for his contributions as a member of my follow-up group, and Assistant Professor Evangelos Karapanos for offering my first research position and teaching me so much in a very short amount of time.

Much of the research presented here and my survival in Finland would have not been possible without the support of my colleagues. In particular, I would like to thank Dr. Simo Hosio, Dr. Marko Jurmu, Dr. Denzil Ferreira, Dr. Hannu Kukka and Mr. Tommi Heikkinen for all their help. Both our serious and ridiculous conversations were important to me over the years.

I thank my pre-examiners Assistant Professor Aniket Kittur from Carnegie Mellon University and Associate Professor Panagiotis Ipeirotis from New York University for their valuable comments and suggestions. In addition, I would like to thank Professor Albrecht Schmidt for agreeing to act as the official opponent in the thesis defense.

I gratefully acknowledge Infotech Oulu as the main financial supporter for this thesis. Further financial support through personal grants and scholarships from the Nokia Foundation, the Walter Ahlström Foundation, the Tauno Tönning Foundation and UniOGS are also gratefully acknowledged

I would also like to thank my friends and family back home. My friends never failed in giving me other things to think about besides work, either by playing games, moaning at each other about the recent football games or simply through conversations. I thank my mom Conceição, my late father João, my sister Izelina and my brother Adelino for all their support and love throughout my life. Finally, I wish to thank my significant other Nanci, for all her support, patience and love throughout this journey.

Oulu, May 2015 Jorge Goncalves

Abbreviations

DRM	Day Reconstruction Method
etc.	et cetera
et al.	et alia
e.g.	exempli gratia
IAPS	International Affective Picture System
i.e.	id est
LAKE	Location Archetype Keyword Extraction
MTurk	Amazon’s Mechanical Turk
NFC	Near Field Communication
QR	Quick Response
R&D	Research and Development
SMS	Short Message Service
SUS	System Usability Scale

Original publications

This thesis is based on the following publications, which are referred throughout the text by their Roman numerals (I–IV):

I Goncalves J, Ferreira D, Hosio S, Liu Y, Rogstadius J, Kukka H & Kostakos V (2013) Crowdsourcing on the Spot: Altruistic Use of Public Displays, Feasibility, Performance, and Behaviours. Proc. The 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013), Zurich, Switzerland: 753–762. DOI: 10.1145/2493432.2493481

II Goncalves J, Hosio S, Ferreira D & Kostakos V (2014) Game of Words: Tagging Places through Crowdsourcing on Public Displays. Proc. The 2014 ACM Conference on Designing Interactive Systems (DIS 2014), Vancouver, BC, Canada: 705–714. DOI: 10.1145/2598510.2598514

III Goncalves J, Pandab P, Ferreira D, Ghahramani M, Zhao G & Kostakos V (2014) Projective Testing of Diurnal Collective Emotion. Proc. The 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2014), Seattle, WA, USA: 487–497. DOI: 10.1145/2632048.2636067

IV Hosio S, Goncalves J, Lehdonvirta V, Ferreira D & Kostakos V (2014) Situated Crowdsourcing Using a Market Model. Proc. The 27th ACM User Interface Software and Technology Symposium (UIST 2014), Honolulu, HI, USA: 55–64. DOI: 10.1145/2642918.2647362

Contents

1 Introduction

"Crowdsourcing has the potential to correct a long-standing human conundrum. The amount of knowledge and talent dispersed among the numerous members of our species has always vastly outstripped our capacity to harness those invaluable quantities. Instead, it withers on the vine for want of an outlet. Crowdsourcing is the mechanism by which such talent and knowledge is matched to those in need of it. It poses a tantalizing question: What if the solutions to our greatest problems weren't waiting to be conceived, but already existed somewhere, just waiting to be found, in the warp and weave of this vibrant human network?."

– Jeff Howe (Howe, 2008)

1.1 Background and motivation

Originally brought into prominence in 2006 by Jeff Howe (2006), crowdsourcing has been adopted as an umbrella term to refer to the coordinated approach in which a computationally challenging task is broken down into several pieces. Those pieces are subsequently distributed and "solved" by humans (typically referred to as "workers"). The division of workload makes this approach highly effective for tasks that can be parallelised. People that participate in these crowdsourcing efforts have different motives for doing so such as payment (Rogstadius *et al.*, 2011), altruism (Goncalves *et al.*, 2013a) or simply contribute without being aware (e.g. reCAPTCHA (Von Ahn *et al.*, 2008)).

The emergence of online crowdsourcing markets (such as Amazon's Mechanical Turk (2014), CrowdFlower (2014), oDesk (2014)) make it convenient to pay for workers willing to solve a range of different tasks. In terms of research, the existence of these crowdsourcing markets makes it possible for researchers to quickly recruit participants or conduct controlled experiments to investigate crowdsourcing itself and what affects task performance.

Online communities like 99Designs (2014), Wikipedia (2014) and Stack Overflow (2014) focus on building and maintaining a community of volunteers (Mamykina *et al.*, 2011), or on enticing experts with high rewards for complex tasks (Heimerl *et al.*, 2012). Some examples of platforms that use the latter strategy to entice experts include Topcoder (2014) that rewards winners with desirable prizes or Innocentive (2014) that requires worker contribution throughout several months to solve R&D challenges.

Further, mobile crowdsourcing systems offer new possibilities for conducting crowd work. They are able to gather real-time information by leveragin the fact that a significant amount of people almost always have their mobile devices with them. This information can then be aggregated and presented in a useful way to other people (e.g. traffic information (Waze, 2015), airport information (GateGuru, 2015)). Furthermore, they are more readily accessible to people in developing countries that might not have access to computers and the Internet enabling them to participate in crowdsourcing efforts (Gupta *et al.*, 2012).

Another approach is to embed input mechanisms (e.g., public displays, tablets) in a physical space and leverage users' serendipitous availability (Müller *et al.*, 2010) or idle time ("cognitive surplus" (Shirky, 2010)). Crowdsourcing under these conditions can be classified as *situated crowdsourcing*. Contrary to mobile crowdsourcing, situated crowdsourcing does not require any deployment effort from workers. Furthermore, it allows for a geofenced and more contextually controlled crowdsourcing environment, thus enabling targeting certain individuals, leveraging people's local knowledge or simply reaching an untapped source of potential workers. Situated crowdsourcing differs from other types of crowdsourcing substantially, offering a complementary – not replacement – means of enabling crowd work.

However, while online and mobile crowdsourcing have been extensively studied, the same cannot be said about situated crowdsourcing. A systematic assessment of the *feasibilit*y and *performance* of situated crowdsourcing is currently lacking, as well as a basic understanding of the *behaviours* of its workers. Therein resides the main focus of this thesis which its articles empirically investigate. Thus, this thesis' research questions are:

RQ1. What considerations should be taken into account when designing crowdsourcing applications for situated technologies?

RQ2. What are the differences between situated crowdsourcing and other means of collecting crowd contributions in terms of task uptake and quality of work?

RQ3. What are the emergent behaviours that crowd workers exhibit when using situated technologies?

1.2 Articles, contribution and author's role

Four articles are included in this thesis, published in relevant international, peer-reviewed conferences in the fields of Human-Computer Interaction and Ubiquitous Computing.

Article I presents the first attempt to investigate altruistic use of interactive public displays in naturalistic usage settings as a crowdsourcing mechanism. The main objective was to compare performance between situated and online crowdsourcing (i.e. Mechanical Turk). The task used was identical across these two platforms: workers were asked to count malaria-infected blood cells on images of a petri dish generated algorithmically. Furthermore, we investigated what effect motivation can have in this setting through the controlled use of motivational design (altruism, adapted community-based and enjoyment-based intrinsic motivators) and validation check mechanisms. Finally, opportunistic interviews and video data allowed us to analyse and better understand worker behaviour when performing situated crowdsourcing tasks. The findings highlight the feasibility of this approach as it can produce a higher rate of task uptake and comparable performance, even to paid studies on Mechanical Turk, given an appropriate motivational design. We also identified several categories of workers that contributed differently in terms of performance, showing that "lone" workers were the more efficient. The thesis author was responsible for designing the study together with the co-authors, implementing the software running on the public displays, deployment, data collection and analysis.

Article II presents a crowdsourcing game for multipurpose public displays (i.e. multiple concurrent services) that enables the creation of a keyword dictionary to describe locations. The main objective was to demonstrate that situated crowdsourcing can effectively leverage people's local knowledge. Furthermore, we investigate the potential of gamification and crowdsourcing to attract workers while using situated technologies as well as filter unwanted contributions through playful crowd-moderation. Our approach had workers contribute to the dictionary in two ways. First, they were given the option to add a keyword of their own to describe their location. Second, they classified as relevant or irrelevant, through gameplay, a randomly selected set of keywords that was already part of the dictionary for that location. The initial worker-generated keywords were then compared to other methods of obtaining these location relevant keywords. These methods were interviews with local citizens, an automated algorithm name LAKE (Kostakos *et al.*, 2013a) and randomly generated words. The findings highlight that our approach can produce more accurate and richer results than the other tested methods while maintaining high levels of participation, even when competing with multiple concurrent services. Additionally, by relying on the crowd to both provide and evaluate input, we demonstrate that despite their public nature, public displays provide reliable results. The thesis author was responsible for designing the study together with the co-authors, deployment, data collection and analysis.

Article III presents a projective test deployed on public displays to study a community's collective diurnal emotion. The main objectives were to demonstrate that situated technologies are ideal to study a community while maintaining a rigorous geographical scope, and that situated crowdsourcing results can be aggregated to provide insights regarding the involved community. The projective test consisted of tagging a person's emotional state (sad, neutral or happy) shown in a video. However, our goal was not the appropriate classification of said videos, but instead to capture the workers' projected emotional state given their errors either in the negative or positive direction. The experiments consisted of three different studies: 1) validation of the generated projective test through priming participants either negatively or positively using the International Affective Picture System (IAPS), 2) unsupervised field deployment in the community we wanted to study over the course of four weeks, and 3) baseline assessment of community emotion using Day Reconstruction Method (DRM) to compare to the results obtained in the second study. The findings demonstrated that our projective test could effectively capture workers' current emotional state and that we successfully crowdsourced the community's diurnal rhythms of emotion consistent with the DRM study, literature on affect and our understanding of the community's daily routine. The thesis author was responsible for designing the study together with the co-authors, implementing the software running on the public displays, deployment, data collection and analysis.

Article IV presents the first situated crowdsourcing platform using a market model. The main objectives were to systematically investigate workers' behaviours and response to economic incentives in a situated crowdsourcing market. For this purpose, we built *Bazaar*, which consisted of a grid of touch-screen kiosks across our campus that rewarded workers with a virtual currency (HexaCoins) for completing tasks. Each worker had their own individual account and could exchange HexaCoins for money, goods (e.g. movie tickets), or with other workers. When workers collected their rewards, we conducted interviews and gathered qualitative insights into their use of the platform. We explicitly decided to provide a variety of crowdsourcing tasks to cater to the varying interests and skills of workers and keep them engaged longitudinally. These tasks included several appealing crowdsourcing applications such as data categorization (counting, identifying), sentiment analysis, content creation, content moderation and a survey. The findings highlight the fact that price mechanism is a very effective tool for adjusting the supply of labour in a crowdsourcing market in terms of location, time of day and task performed. Furthermore, we demonstrated that situated crowdsourcing can attract a populous workforce with comparable quality of work to its online counterparts while maintaining a higher task uptake and reaching untapped populations of workers. *Bazaar* also enabled us to track individual worker behaviour and in conjunction with our qualitative data we were able to provide a holistic assessment of worker behaviour in a situated crowdsourcing market. Finally, virtual currency was shown to be a key enabler to sustaining workers' engagement and interest in our platform. The thesis author was responsible for designing the study together with the co-authors, assisting the implementation and deployment of the software running on the tablets, data collection and analysis.

1.3 Thesis outline

The remainder of the thesis is organized as follows: in Chapter 2, we introduce the related work. We start by defining situated technologies and why their affordances make them an appropriate crowdsourcing deployment environment. This is then followed by a look at relevant literature on crowdsourcing beyond the desktop, focusing on work done with mobile phones and the aforementioned situated technologies. Next, we look at the crucial issue of quality control focusing on its two overarching categories: appropriate task design and post-hoc analysis of the results. We conclude this chapter with previous work on the role of motivation and rewards in online crowdsourcing markets followed by a reflection on the challenges of this issue when using situated technologies.

Chapter 3 discusses the experiments and contributions of the four articles included in this thesis. This chapter is divided in three main themes based on our research questions: appropriate crowdsourcing task design for situated technologies (RQ1), comparison of performance between situated crowdsourcing and other means of collecting crowd contributions (RQ2), and the emergent worker behaviours of situated crowdsourcing (RQ3). We end this chapter by providing a summary of the contributions. In Chapter 4, the original research of this thesis is discussed in respect to the related work, and we revisit the research questions put forth in the beginning of the thesis. Chapter 5 concludes the thesis.

2 Related Work

In this chapter, we define the deployment environment prevalent in all articles of the thesis: situated technologies in public spaces. We then present an overview of relevant work on crowdsourcing beyond the desktop, followed by the important challenges of quality control and motivational aspects in crowdsourcing.

2.1 Situated technologies

We begin by investigating what makes situated technologies (e.g. public displays and embedded tablets) an attractive medium for the deployment of crowdsourcing projects. We do so by discussing some of their key affordances, but also some of their more problematic features.

Situated technologies are computational elements in everyday environments (Greenfield & Shepard, 2007). As they become increasingly affordable, researchers have systematically attempted to identify novel applications like hedonic services (i.e. games, opinion disclosure) or information-based services (i.e. information boards) that offer instant benefits to users (Brignull & Rogers, 2003; Churchill *et al.*, 2004; Hosio *et al.*, 2012; Schroeter *et al.*, 2012). One of their characteristics is that their use is not motive-driven but serendipitous (Müller *et al.*, 2010) making them viable opportunistic contribution mediums. This allows them to reach a larger audience than conventional means (e.g. paper forms) (Hosio *et al.*, 2014a; Hosio *et al.*, 2015), and the mere act of using them attracts more people (Brignull & Rogers, 2003; Kukka *et al.*, 2013; Goncalves *et al.*, 2014d). Furthermore, since they can be deployed in a particular place, they can effectively target certain individuals and also leverage local knowledge of people within that place. These affordances enable situated crowdsourcing to potentially overcome some of the limitations of online and mobile crowdsourcing.

Despite the various benefits of situated technologies, there are some serious challenges. For instance, it can be difficult to ensure credibility (Taylor *et al.*, 2012) if users do not feel that a deployment is legitimate. This in turn can lead to a decrease in perceived efficacy and, more importantly, users can feel discomfort when using technology in public spaces as interacting with them may make users the centre of everyone's attention (Kuikkaniemi *et al.*, 2011). Previous work addressed this issue through a playful situated feedback application connected to social network services on public displays called *Ubinion* (Hosio *et al.*, 2012). Studies with the *Ubinion* prototype also highlighted situated public displays being fit for acquiring contextually relevant feedback. Similar projects developed feedback systems for campus settings, utilising online interfaces, dedicated mobile clients, and Twitter as inputs (Day *et al.*, 2007; Munson *et al.*, 2011). Another challenge of some situated technologies (e.g. public displays) is that they can have less rich interface controls, when compared to standard desktop environment, which can result in limited usability and accessibility. This means that not all types of contributions or tasks can be crowdsourced on these technologies. Moreover, in-the-wild deployments that use situated technologies typically do not track specific individuals and it can become challenging to interpret gathered results. Finally, their maintenance is more difficult than, for example, maintaining an online server and can incur higher initial costs.

2.2 Crowdsourcing beyond the desktop

In this section, we discuss literature on crowdsourcing using mobile phones and situated technologies. We also discuss more broad applications of situated technologies where people's contributions are collected.

2.2.1 Mobile crowdsourcing

Mobile crowdsourcing has allowed researchers to push tasks to the workers, anywhere and anytime (Goncalves *et al.*, 2014c). Targeting low-end mobile phones, *txtEagle* (Eagle, 2009) is a platform for crowdsourcing tasks specific to inhabitants of developing countries. Similar platforms are *MobileWorks* (Narula *et al.*, 2011) and *mClerk* (Gupta *et al.*, 2012) that specifically focus on asking workers to convert handwritten words to typed text from a variety of vestigial dialects. In a larger project, a mobile crowdsourcing platform called *MoneyBee* (Govindaraj *et al.*, 2011) was made accessible to mobile phone users in emerging markets through their mobile operators resulting in a higher number of potential workers.

Targeting smartphones, Alt *et al.* (2010) explore location-based crowdsourcing for distributing tasks to workers. They focus on how workers may actively perform real-world tasks for others, such as giving a real-time recommendation for a restaurant, or providing an instant weather report wherever they are. Similarly, Väätäjä *et al.* (2011) report a location-aware crowdsourcing platform for authoring news articles by requesting photographs or videos of certain events from its workers. Mashhadi & Capra (2011) suggest using contextual information, such as mobility, as a mechanism to ensure the quality of crowdsourced work. *Wallah*, a mobile crowdsourcing platform for Android OS, implements caching for offline situations and aims to minimize the impact of different screen sizes of smart phones (Kumar *et al.*, 2014). Finally, *mCrowd* (Yan *et al.*, 2009) enables mobile users to utilise sensors on their smartphone to participate and accomplish crowdsourcing tasks, including geolocation-aware image collection, image tagging, road traffic monitoring, and others.

Recently, an active community has grown around the topic of crowdsourcing measurements and sensing. This participatory sensing movement is part of the larger concept of "Citizen Science" (Paulos *et al.*, 2008) that relies on mobilising large parts of the population to contribute to scientific challenges via crowdsourcing. Often, this involves the use of smartphones for collecting data (Aoki *et al.*, 2008; Burke *et al.*, 2006) or even donating computational resources while the phone is idle (Arslan *et al.*, 2012). In addition, there are several platforms that enable the rapid collection of data from the crowd through their phones and other multiple sources (e.g. social media). For example, *Ushahidi* (2014) is an open-source crisis map platform created in 2007 and has been deployed in several problematic areas like Haiti, Kenya and Afghanistan. A leisure-oriented platform called *Clickworker* (2013) enables the crowd to gather, verify and research local information like meal deals, city trips or restaurant reviews using their smartphones.

Despite the appeal of mobile phones, using them for crowdsourcing requires workers' *explicit* deployment, configuration and use of the device. For example, in SMS-based crowdsourcing, participants need to explicitly sign up for the service, potentially at the cost of a text message exchange. This challenges recruitment of workers, as a number of steps normally needs to be performed before a worker can actually start contributing using their device. Furthermore, workers normally cannot control when they receive requests on their phones unless the system allows them to specify when they wish not to be disturbed (Church *et al.*, 2014).

2.2.2 Collecting crowd contributions through situated technologies

There has been very little research conducted using situated technologies for crowdsourcing. In one of the few examples, Heimerl *et al.* reported *Umati* (2012), which used a vending machine with a touch display for locally relevant tasks. Workers could earn credits performing these tasks and exchange them for physical rewards directly on the vending machine (snacks). In a position paper, Marshall *et al.* (2011) outline a project, where its goals were to reduce staff stress levels and increase restorative opportunities through situated crowdsourcing. Unfortunately, to the best of our knowledge, this project never left the planning phase.

However, because situated technologies are typically available to everyone within a location, they have been explored for democratic participation and civic engagement to find novel ways of improving the efficacy of citizens to contribute to their local communities. For example, *Viewpoint* allowed local organizations to post questions in public spaces, and citizens to vote (Taylor *et al.*, 2012). Similarly, *VoiceYourView* engaged people to voice out their opinions about the design of the surrounding space, utilizing natural language processing and speech recognition to collect real-time feedback (Whittle *et al.*, 2010). Its clever design of a traditional-looking telephone as the input mechanism is a great example of how deployments in the public space can quickly attract the interest of hundreds with barely any promotion. Another example is *PosterVote* (Vlachokyriakos *et al.*, 2014), which successfully used paper posters and a lightweight hardware to enable and support local political activism. Fortin *et al.* discussed a media façade installation called *Mégaphone* (Fortin *et al.*, 2014), intended to act as a digital "speakers' corner." The system consisted of a public microphone on a stand with speech recognition, loudspeakers and two media façades. During the deployment, keywords were extracted from the speech and projected to the media façades. This way, people could both contribute to the system through the microphone, as well as consume media from the system through the loudspeakers, as well as the keywords from the media façades. Their findings showed that appropriation of the system was linked to identity-building and place-making while also facilitating social interactions around the installation area.

Despite these innovative feedback collections mechanisms, most archetypal feedback applications on situated technologies utilise typing in some form as their main input modality. Brignull & Rogers reported *Opinionizer* (2003), a system that combined a projected screen with a laptop to type feedback and converse about the everyday contexts it was deployed in. They noted the honey-pot effect and emphasized social pressure and awkwardness that users often feel when interacting publicly. However, typing on public displays has also been shown to be problematic, leading to misuse (Goncalves *et al.*, 2014b). Ananny & Strohecker leveraged public screens and SMS to create public opinion forums (2009). Their *TexTales* installations highlighted how urban spaces can become sites for collective expression and nurture informal, often amusing discussions among its inhabitants. Similarly, Schroeter *et al.* presented a real-world installation for a public forum application called *Discussions in Space* (2012). Their installation was geared to engage passers-by to submit locative and topical messages through SMS and Twitter to a large display, in order to facilitate public deliberation of issues topical to the actual location. Their framework for engagement includes three factors, namely people, content and location. For a public display to serve its purpose effectively, a sweet spot should be identified, where the deployed display effectively fulfils each of the three factors. Other alternative approaches include simplified input mechanisms like the one presented by the *Vote with Your Feet* prototype in which users voted on local public pools using their feet (Steinberger *et al.*, 2014) or moving the input to a handheld personal device (Boring *et al.*, 2009; Hosio *et al.*, 2014a).

2.3 Quality control in crowdsourcing

A crucial element of crowdsourcing, due to its susceptibility to gaming behaviour, is quality control. Gaming behaviour is usually exhibited by workers in an attempt to reap rewards with minimum effort (Downs *et al.*, 2010). Two quality control approaches are appropriate task design and post-hoc analysis of results (Kittur *et al.*, 2013), both of which were used and adapted in our research in order to address our first two research questions.

2.3.1 Appropriate task design

Crowdsourcing performance can be substantially improved by taking a number of steps when designing the task itself. For instance, previous work has suggested the inclusion of explicitly verifiable questions (Kittur *et al.*, 2008) when developing the experiment. The inclusion of these “fact-checking” questions has been shown to improve the quality of completed tasks as workers become aware of prompt response verification. Furthermore, the actual difficulty of the task has an effect on task performance (Rogstadius *et al.*, 2011). Specifically, for tasks with higher difficulty levels, workers may simply give up, or provide an approximate answer for the task. Therefore, less difficult tasks are more likely to be completed (Horton *et al.*, 2011).

Agreement filters have also been shown to be an effective quality control mechanism. Von Ahn & Dabbish (2004) leveraged this method with their *ESP* game which could be played by two partners that would see the same image and then have to provide the same answer before proceeding to the next one. In another example, a document writing and editing prototype named *Soylent* (Bernstein *et al.*, 2010) proposed the Find-Fix-Verify pattern in which its Verify phase randomised the order of the unique alternatives generated in the Fix phase and asked several workers to vote on them. Similarly, Dow *et al.* (2012) used peer-review in their *Shepherd* prototype by providing workers with timely and expert feedback when writing consumer reviews.

Finally, splitting up the work into fault-tolerant subtasks has also been shown to improve worker performance. For example, both *Soylent (Bernstein et al., 2010)* and *PlateMate* (Noronha *et al.*, 2011) had an input decomposition phase at the start. In a more high-level approach, Kittur *et al.* (2011) proposed a general purpose framework named *CrowdForge* for distributed processing and providing scaffolding for complex crowdsourcing tasks. While appropriate upfront task design can significantly improve the quality of the work, often it is still necessary to analyse collected data to further improve its reliability and validity, as we discuss next.

2.3.2 Post-hoc analysis of results

Another approach for quality control is post-hoc filtering of substandard contributions. One of the most commonly used techniques is to adopt a Gold Standard (Downs *et al.*, 2010). This entails the creation and inclusion of tasks that have known answers to the requested crowdsourcing job. The inclusion of these pre-labelled questions can enable a task requester to capture the reliability of its workers as they are unware which particular tasks have this feature. Therefore, answers from workers that performed badly on the pre-labelled tasks can be removed, potentially improving the accuracy of the crowdsourced results. However, authoring gold data can be burdensome, and gold standards are challenging to implement in subjective or generative tasks, like writing an essay (Kittur *et al.*, 2013).

Another approach is to analyse the extent to which workers agree with each other in their answers (Callison-Burch, 2009; Ipeirotis *et al.*, 2010). High agreement between workers increases reliability and can even allow ranking of contributions when the task does not have a single correct answer (Goncalves *et al.*, 2014a). A troublesome characteristic of this method is that it is susceptible to collusion (Douceur, 2002) which can be exacerbated by the creation of fake accounts in an attempt to thwart quality assurance methods (Kittur *et al.*, 2013).

Finally, a more invasive approach has been proposed that entails examining the way workers perform the work instead of their contributions. For instance, *CrowdScape* (Rzeszotarski & Kittur, 2012) enabled worker evaluation through interactive visualisation and mixed initiative machine learning, while another prototype used a task fingerprinting system to monitor worker activity on crowdsourcing markets (Rzeszotarski & Kittur, 2011). A simpler approach to this method entails requesters reviewing work-in-progress snapshots of the worker's desktop, a possibility in the oDesk's (2014) crowdsourcing platform.

2.4 Motivation and rewards

Studies on crowdsourcing have shown that workers often require some kind of motivation or rewards to participate and effectively spend time doing crowd work. In other words, crowdsourcing performance is not just a matter of channel or medium, but also a matter of motivation (Kaufmann *et al.*, 2011) and rewards (Kittur *et al.*, 2008). Kaufmann *et al.* (2011) identify two major types of motivation for crowdsourcing: intrinsic and extrinsic. They note that intrinsic motivation can be enjoyment-based (related to the fun and enjoyment that the contributor experiences through their participation) and community-based (related to community participation, and include community identification and social contact). Extrinsic motivation can relate to having immediate or delayed payoffs or rewards (including material benefits), as well as social motivation (such as values and beliefs). Next, we look more concretely at motivational approaches and rewards of crowdsourcing in online markets and the challenges regarding situated technologies.

2.4.1 Motivation in online markets

As online markets constitute the majority of available crowdsourcing tasks, it is important to provide an overview of why people take part as workers in crowdsourcing markets, and what does the theory suggest about their performance in completing tasks. A traditional "rational" economic approach to eliciting higher quality work is to increase extrinsic motivation (Gibbons, 1997). In other words, an employer or task requester can increase how much they pay for the completion of a task to improve the quality of submitted work. Some evidence from traditional labour markets supports this view: Lazear (2000) found workers to be more productive when they switched from being paid by time to being paid by piece; Hubbard & Palia (1995) found correlations between executive pay and firm performance when markets were allowed to self-regulate.

An experiment by Deci (1975) found a "crowding out" effect of external motivation, such that students paid to play with a puzzle later played with it less and reported less interest than those who were not paid to do so. In the workplace, performance-based rewards can be "alienating" and "dehumanizing" (Etzioni, 1971). If the reward is not substantial, then performance is likely to be worse than when no reward is offered at all; insufficient monetary rewards can act as a small extrinsic motivation that tends to override the possibly larger effect of the task's likely intrinsic motivation (Gneezy & Rustichini, 2000). Given that crowdsourcing markets such as Mechanical Turk tend to pay very little money and involve relatively low wages (Paolacci *et al.*, 2010), external motivations such as increased pay may have less effect than requesters may desire. Indeed, research examining the link between financial incentives and performance in Mechanical Turk has generally found a lack of increased quality in worker output (Mason & Watts, 2009). The relationship between price and quality has also had conflicting results in other crowdsourcing applications such as answer markets (e.g. (Harper *et al.*, 2008)). Although paying more can get work completed faster, it has not been shown to get work done better.

Another approach to getting work done better could be increasing the intrinsic motivation of the task. Under this view, if workers find the task more engaging, interesting, or worth doing in its own right, they may produce higher quality results. Unfortunately, evidence so far has not fully supported this hypothesis. For example, while crowdsourcing tasks framed in a meaningful context motivate individuals to do more, they are no more accurate (Chandler & Kapelner, 2013). On the other hand, work by (Rogstadius *et al.*, 2011) suggests that intrinsic motivation has a significant effect on workers' performance.

These contradictory results and a number of other issues that suggest the question of motivating crowd workers has not yet been definitively settled. First, prior studies have methodological problems with self-selection, since workers may see equivalent tasks with different base payment or bonuses being posted either in parallel or serially. Second, very few studies besides (Rogstadius *et al.*, 2011; Shaw *et al.*, 2011) have looked at the interaction between intrinsic and extrinsic motivations; Mason & Watts (2009) vary financial reward (extrinsic), while Chandler & Kapelner (2013) vary meaningfulness of context (intrinsic) in a fixed diminishing financial reward structure. Finally, the task used in Chandler & Kapelner (2013) resulted in very high performance levels, suggesting a possible ceiling effect on the influence of intrinsic motivation.

2.4.2 Motivation in situated crowdsourcing

The increased ease, from the part of the worker, with which it is possible to participate in crowdsourcing work through situated technologies, raises new possibilities. However, it can also raise important questions of motivation and performance: if people can contribute to crowdsourcing using a non-personal device in public, *why would they choose to do so and what will be the quality of the work*?

While the issue of motivation has been a long-standing concern in the design of computer systems and online services, new technologies and contexts require that *new motivational approaches are developed, adapted,* and *validated*. In terms of crowdsourcing, research in psychology, sociology, management and marketing provide a solid theoretical basis on human motivation (Kittur *et al.*, 2013). However, these theoretical approaches typically have to be adapted and fine-tuned for a crowdsourcing setting, even more so when it comes to situated crowdsourcing. At the same time, by motivating workers to contribute more, task requesters can unwillingly make them more susceptible to quality control issues (Kittur *et al.*, 2013), so careful motivational considerations have to be taken into account.

In addition to accounting for human behaviour, a motivational approach also needs to account for the technologies and context of use. While situated technologies can offer timely contextual information, it becomes increasingly difficult to maintain people's interest and engagement in these environments (Goncalves *et al.*, 2013b). A further challenge with situated technologies is that it is typically in the hands of users, away from the control of a lab setting, and may produce "noisy" results due to unpredictable behaviour from users (Hosio *et al.*, 2014a; Schroeter *et al.*, 2012). Thus, prior work further emphasises the importance of appropriate motivational approaches to address these challenges. For instance, Heimerl *et al.* used a touch-screen vending machine to attract workers and reward them with snacks (2012).

The articles in this thesis adapt, present and validate motivational approaches for *situated crowdsourcing.* These approaches draw on prior literature in human behaviour, account for situated technologies, and are *validated* in field trials to establish their effect on *task uptake* and *quality* of work.

3 Research contributions

This chapter presents the research contributions of Articles I-IV that are relevant to this thesis and its research questions. We start by explaining the design decisions and their impact on our deployments. We then look at the performance of our situated crowdsourcing applications in terms of task uptake and accuracy. We also provide a summarised performance comparison between our deployments and previous crowdsourcing work using different technologies (online, mobile and situated). Finally, we discuss the emergent worker behaviour in situated crowdsourcing. Table 1 summarises for each of the articles: the research objectives, the experimental setup and the findings that are relevant to this thesis.

Table 1. Summary of research contributions.

Article	Research objectives	Experimental setup	Contributions
I	• Investigate the influence of intrinsic motivators in situated crowdsourcing performance. • Compare performance between situated and online crowdsourcing. • Investigate workers' social dynamics around situated technologies.	Twenty-five days deployment on four single-purpose displays.	• Intrinsic motivators and fact-checking questions improved the quality of work in situated crowdsourcing (RQ1). • Situated deployment was shown to have a higher task uptake than the online deployment with comparable quality of work (RQ2). • Six different behaviours of workers were identified (RQ3).
II	• Investigate the potential of gamification to motivate workers to perform well in situated crowdsourcing. • Create a keyword dictionary that is relevant to the location where a situated technology is deployed by leveraging local knowledge. • Explore relying on workers to rank contributions, and filter out unwanted ones in situated crowdsourcing.	One month deployment on five multipurpose displays.	• Gamification successfully motivated workers to perform high quality tasks in situated crowdsourcing (RQ1). • By using the crowd to both provide and evaluate input, we demonstrate that despite their public nature, situated technologies can provide reliable results (RQ1). • Situated crowdsourcing approach produced more accurate and richer results than the other tested methods even when competing with multiple concurrent services (RQ2).
III	• Develop and validate a projective test to detect individuals' current emotion. • Use the projective test to provide a community's diurnal collective emotion through situated crowdsourcing. • Compare the developed method to Day Reconstruction Method (DRM) results.	Four weeks deployment on four single-purpose displays.	• A simplified and more engaging task design led to less curiosity clicks and more tasks completed per session when compared to Article I (RQ1). • The aggregated crowdsourcing results successfully captured the community's diurnal rhythms of emotion consistent with the DRM study, literature on affect and our understanding of the community's daily routine (RQ2).

Article	Research objectives	Experimental setup	Contributions
IV	• Investigate the feasibility of a situated crowdsourcing market with a variety of tasks and worker payment. • Determine if it is possible to control labour supply in a situated crowdsourcing setting. • Track situated crowdsourcing workers to identify emergent worker behaviours.	Three week deployment on four public tablets.	• A price mechanism was shown to be a very effective tool for adjusting the supply of labour in a situated crowdsourcing market (RQ1). • Price-setting should follow contextual and cultural factor and not blindly follow online prices (RQ1). • A situated crowdsourcing market can attract a populous workforce with comparable quality of work quality to its online and mobile counterparts while maintaining a higher task uptake (RQ2). • An assessment of worker behaviour in a situated crowdsourcing market was provided (RQ3).

3.1 Designing situated crowdsourcing applications

The design of situated crowdsourcing applications requires careful consideration, as one cannot simply adopt the task design from online markets. In the articles present in this thesis, we adapted motivational approaches to better fit this context, as well as tweaking the features of the task itself.

3.1.1 Design considerations in Article I

In Article I, the main objective was to compare performance between situated and online crowdsourcing (i.e. Mechanical Turk). It constituted the first attempt to investigate altruistic use of interactive public displays in natural usage settings as a crowdsourcing mechanism. Our crowdsourcing task was deployed on four single-purpose public displays (Figure 1). We relied on workers' altruistic behaviour and appeal to their innate desire to "do something good." However, we decided that since our chosen task (counting malaria cells) is not a local issue, we still needed to have an additional motivation strategy. Hence, we used two types of intrinsic motivation proposed by Kaufman *et al.* (2011): *enjoyment-based* and *community-based*. We used only one construct per motivator, to enable reliable testing on the public display. These constructs were *task identity* and *community identity* as we deemed them to be the more appropriate, given the deployment context. The use of these motivators also meant that we were able to increase the overall credibility of our crowdsourcing application, a common problem in situated technologies deployments (Taylor *et al.*, 2012). In addition, we included a "fact-checking" question before any task could be completed. We argue that this approach is better suited for situated crowdsourcing than using gold standards. Besides avoiding gold standard limitations, it also tackles head-on the potential lack of seriousness that is normally part of in-the-wild deployments (Brown *et al.*, 2011), as workers become aware of prompt response verification.

Fig. 1. Single-purpose displays used in Article I and III.

Before our field deployment, we also had a pilot study in which we initially used all images from Rogstadius *et al.* (2011). However, it became quickly apparent that images of higher complexity were too time-consuming and challenging to be completed on a public display. Thus, we opted to only use a subset of 30 images with the lowest complexity (Figure 2). This suggests that more complex tasks may not be ideal for situated technologies. The overall complexity of the task may also account for the high number of *curiosity clicks* as only 27% of people that interacted with the display completed at least one task, and completed on average only 2.5 tasks each.

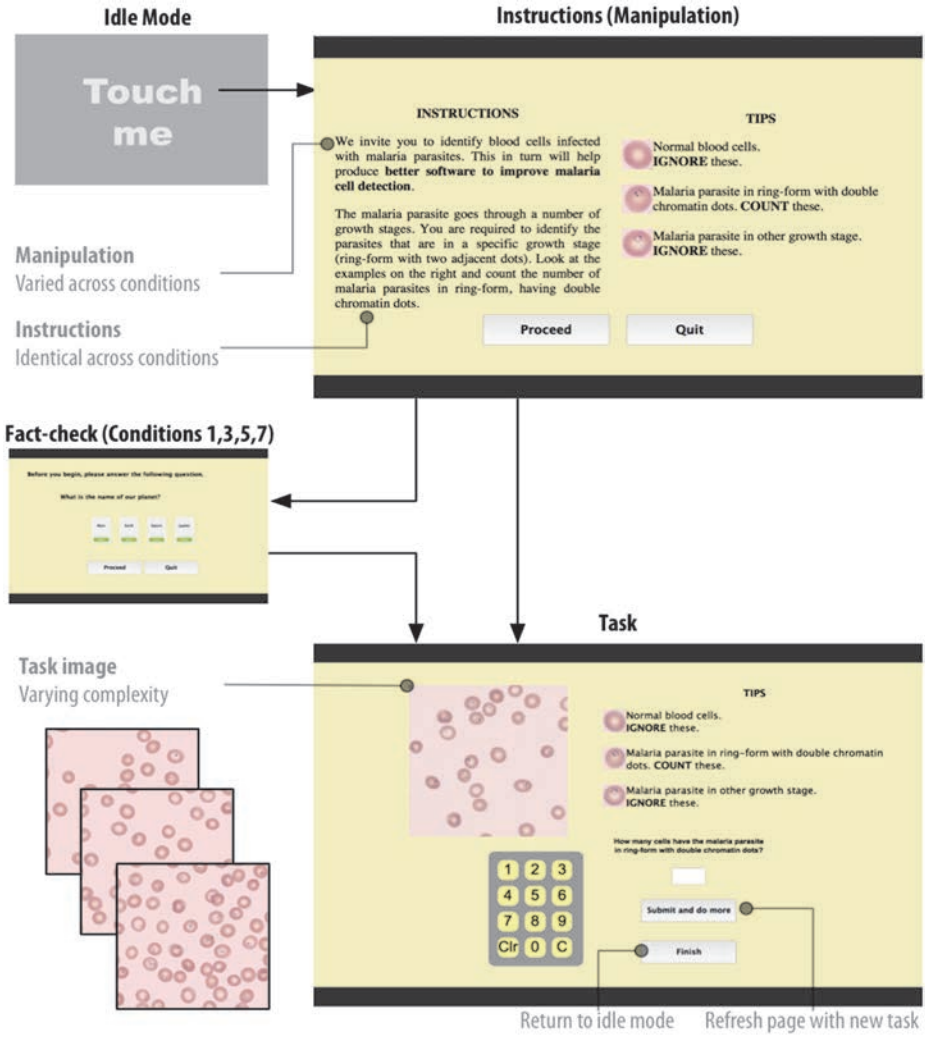

Fig. 2. Screenshot flowchart of the experimental task reported in Article I. A worker was allocated to a condition when clicking the idle mode page. Only half of the conditions had the fact-check page. Manipulation text varied across condition and the task image varied in complexity.

3.1.2 Design considerations in Article II

In Article II, the main objective was to demonstrate that situated crowdsourcing can effectively leverage people's local knowledge. To this end, we created a crowdsourcing game for multipurpose public displays (i.e. multiple concurrent services) that enabled the creation of a keyword dictionary to describe locations. Hence, the deployment environment changed slightly in that we no longer used single purpose displays, where our crowdsourcing application would always be in the foreground. This means that our crowdsourcing application had to compete with dozens of other services (Hosio *et al.*, 2013a; Kostakos *et al.*, 2013b) for the user's attention. Thus, we decided to use gamification as a motivator through gameplay elements like a leaderboard (Figure 3). Games are in general appealing to public display users and are often cited as "unexpectedly popular" (Ojala *et al.*, 2012), implying that people like to use public displays and other situated technologies in a casual way, to spend free time.

Fig. 3. Multipurpose public displays used in Article II.

As for the design of the actual game, we implemented two features: the ability to add new keywords and voting on keywords that are currently part of the dictionary (Figure 4). In order to reduce as much as possible the barriers to contribution, we decided not to make the first feature mandatory as it could deter participants from playing the game as typing on public displays has been shown to be cumbersome (Goncalves *et al.*, 2014b). As for the second feature, we spent considerable effort and time optimising the gameplay settings of the game. We followed the findings in Article I which suggest that crowdsourcing tasks on public displays should be kept short, otherwise the workers are likely to abandon the task and walk away. More importantly, we wanted to avoid the cut-off point in terms of task difficulty, beyond which error rates increase and completion rate dramatically drops, as seen in Article I. In addition, because the scope of the crowdsourcing task was highly localised, we had to conceive other means of evaluating the quality of the crowdsourced keywords as simply deploying the same task online was not an option. To this end, we bootstrapped the game by populating each display's dictionary with keywords from different sources: 1) keywords generated algorithmically through LAKE analysis (Kostakos *et al.*, 2013a), 2) keywords obtained by interviewing local citizens and 3) random keywords. This also allowed us to compare the quality of the crowd-generated keywords to the ones generated utilising these methods.

Finally, we purposely kept moderation light for the added keywords meaning that we only removed particularly offensive words. We did this so that the game kept being challenging throughout the deployment: if all the words that appeared were clearly relevant, the gameplay would require little effort and become boring. This means that the "noise", i.e. irrelevant keywords added by workers, contributed to the gameplay experience positively by increasing its difficulty. In contrast with the deployment presented in Article I, workers did not completed the task in an altruistic manner, they simply wanted to pass time and be entertained (Von Ahn & Dabbish, 2008) for which challenge is a key aspect of a game (Sweetser & Wyeth, 2005).

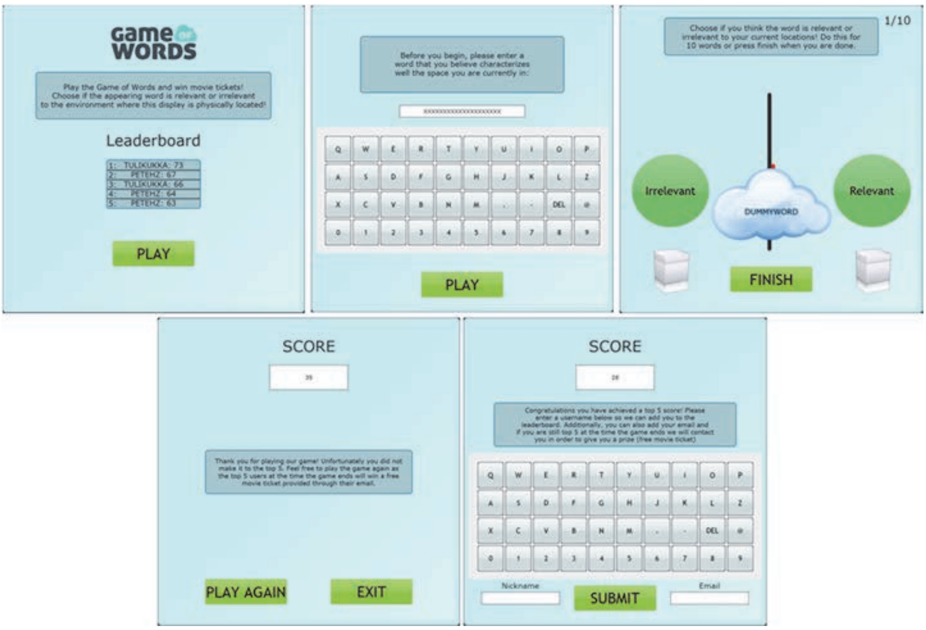

Fig. 4. Different screens of the game reported in Article II.

3.1.3 Design considerations in Article III

In Article III, the main objective was to demonstrate that situated technologies are ideal to study a community while maintaining a rigorous geographical scope, and that situated crowdsourcing results can be aggregated to provide insights regarding the involved community. Our use case consisted of measuring a community's diurnal collective emotion through a crowdsourcing application on public displays based on workers subjective judgement. Subjective judgments are a commonly used information source for scientific experiments. Unfortunately, subjective judgments tend to be problematic because there are no public criteria for assessing judgmental truthfulness (Prelec, 2004). In order to minimise this limitation and possible self-selection bias from participants, we ultimately decided to develop a projective test (Figure 5). These tests are designed using ambiguous stimuli, presumably revealing hidden emotions projected by the person into the test (Exner & Erdberg, 2005). Based on the literature on emotion, the chosen stimulus was the silhouette of a walking person as humans have an innate ability to understand emotions expressed through posture and gait (Atkinson *et al.*, 2007). These silhouettes were extracted from a real-world setting (i.e. after being given a verdict in a Television talent show) where individuals express their emotions genuinely unlike many other available datasets. The task did not require typing and was ambiguous enough to allow individuals to project their emotions while ultimate having a correct answer. We were interested in the mistakes made by the workers, in the positive or negative direction, when identifying the emotion exhibited by the person in the video. Controlled validation of our developed projective test demonstrated that it could effectively gather individuals' emotional state.

Moreover, it was imperative to have enough ratings overall and also in any given hour throughout the work day to enable reliable diurnal aggregate emotion detection. Our crowdsourcing projective test was designed to tackle two limitations of the study presented in Article I as the deployment environment was exactly the same (identical single purpose displays and locations used, Figure 1). These limitations were the high number of curiosity clicks (73% of all interactions), and the low number of tasks completed in each session (2.5). Our first design decision was towards making the crowdsourcing task simpler. This meant simplifying the input mechanism, and therefore lower the time required to answer each task. In addition, by using videos (i.e. dynamic stimuli) instead of static images, our task became more engaging (Lewalter, 2003) and efficient at grabbing people's attention in a public setting. This meant that a higher number of workers would move beyond curiosity clicks and actually contribute with votes which ultimately led to a much higher percentage of workers doing so (90% vs. 27%). Furthermore, these workers also ended up completing more tasks when at the display (6.8 vs. 2.5).

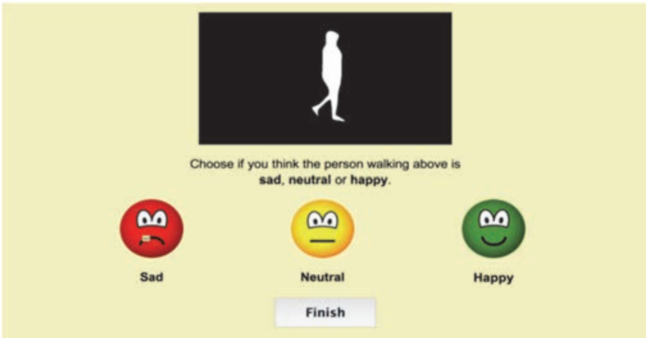

Fig. 5. Task interface of the deployment reported in Article III.

3.1.4 Design considerations in Article IV

Finally, the study described in Article IV explored the feasibility of a situated crowdsourcing market instead of single one-off tasks. The main objectives were to systematically investigate workers' behaviours and response to economic incentives in a situated crowdsourcing market. For this purpose, we built *Bazaar*, which consisted of a grid of touch-screen kiosks across our campus that rewarded workers with a virtual currency (HexaCoins) for completing tasks. The use of a market approach constituted a significant shift from our previous deployments, and we had to rethink our approach regarding the design of the situated crowdsourcing application.

To improve sustainability, we provided a variety of crowdsourcing tasks (Figure 6) to cater to the varying interests and skills of workers (Kittur *et al.*, 2013) and to keep them engaged longitudinally. We selected a diverse set of tasks with different stimuli (text, images, videos, current context) and purposes (data categorization: counting and identification, sentiment analysis, content creation, content moderation, survey). In addition, drawing from economic theory (Lehdonvirta & Castronova, 2014) and practical experience, we decided to use a market-driven model (rather than intrinsic motivation) which can enhance its sustainability, and used a virtual currency (rather than directly goods) to provide enhancements to the market's operation, which was redeemable for goods. In order for this to be possible, we had to implement a mechanism to identify workers and store their virtual currency. The initial screen of *Bazaar* allowed workers to register or login to the system (Figure 6). Registration required just a username and password because a lengthy process can reduce participation, especially when using situated technologies (Brignull & Rogers, 2003).

Worker payment in *Bazaar* also meant that this was the only deployment out of the four presented in this thesis that extrinsically rewarded workers. Rewarding workers in situated crowdsourcing remains a relatively unexplored issue. We adopted a price-setting that took into account 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 situated technologies in question. So while our reported hourly rate of roughly 10€ can be deemed too high and a significant factor in the use of the system, we argue that this is not the case. 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 (MTurk average salary per hour is \$5-\$7) in cheap labour countries where many workers on several online marketplaces come from.

Furthermore, the prototype was deployed on public tablets (Figure 6) instead of larger public displays like the deployments reported in the other articles. A disadvantage is that these smaller kiosks would not be as good at attracting workers as their larger display counterparts. However, with *Bazaar*, workers had access to an input device with greater usability, which meant we could include more cumbersome tasks like textual content creation and lead to overall less fatigue.

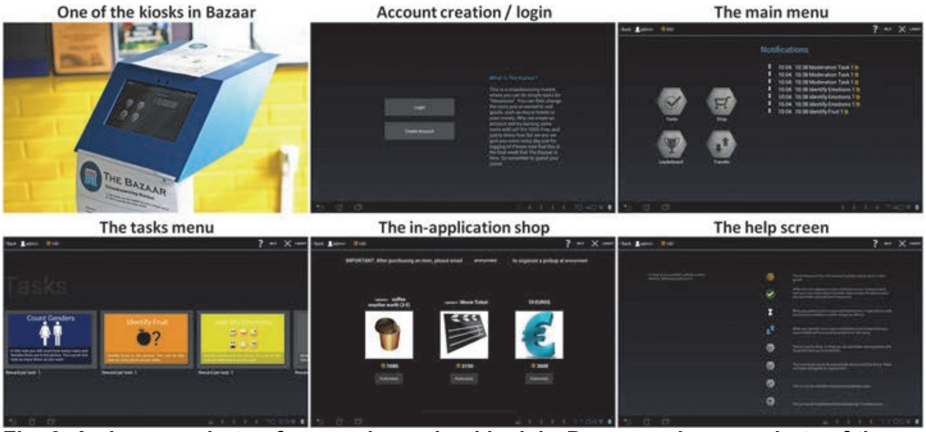

Fig. 6. A close up photo of a crowdsourcing kiosk in *Bazaar* and screenshots of the application screens in Article IV.

In conclusion, throughout our situated crowdsourcing deployments we carefully designed our tasks to appropriately fit the context of use and the identified research objectives. The design considerations reported were ultimately factors that contributed to the improvement of performance in our deployments which we discuss next.

3.2 Situated crowdsourcing performance

In our work, we provide a systematic assessment of performance in situated crowdsourcing. Below, we look at task uptake and quality of work in the deployments reported in this thesis' articles. We conclude this section by providing a comparison between the worker performance in our presented articles and a few other research projects in the literature that used different types of crowdsourcing (Table 2).

3.2.1 Attractiveness and task uptake

A key affordance of situated technologies is that they are effective at attracting users (Kukka *et al.*, 2013). With this in mind, our deployments had little or no promotion, relying mainly on the serendipitous nature and attractiveness of situated technologies (Kukka *et al.*, 2013; Müller *et al.*, 2010). Articles I and III used a simple idle mode initial screen that said "Touch me" (Figure 1) to entice interaction (Kukka *et al.*, 2013). The kiosks used for Article IV had an A3-sized poster attached to them giving more details about the platform. As a result of this affordance and the careful design considerations mentioned previously, our deployments had relatively high levels of task uptake when compared to online and mobile crowdsourcing projects (Table 2).

In Article I, we compare the results of a task deployed on both a public display and on Mechanical Turk. No payment was given to the participants who used the public displays, while participants on Mechanical Turk were rewarded with 0, 3 or 10 cents per the task completed. A direct comparison between the non-paid conditions shows that the rate of uptake on the public display was much higher than on Mechanical Turk, reaching the set objective of 1,200 tasks in 25 days (48 per day), while on Mechanical Turk, 100 tasks were complete in 48 days (2.1 tasks per day). Our public display deployment also outperformed the paid conditions on Mechanical Turk in terms of task uptake: the 3 cents condition had 200 tasks completed in 18 days (11.1 per day), while the 5 cents conditions completed the same amount of tasks in 13 days (15.4 per day). Furthermore, workers in Article I exposed to our motivators completed more tasks per session, further highlighting the importance of motivation on task uptake.

Similarly, the crowdsourcing game presented in Article II was played 632 times generating 362 keywords and 6,009 votes in 30 days (21.1 games, 12.1 keywords and 200.3 votes, per day). Again, thanks to our careful design considerations, the vast majority of players completed each game session in its entirety voting on all ten keywords presented. On average, there were 9.51 votes in each game session. Also, the amount of games played, and keywords and votes collected, remained constant throughout the deployment suggesting the medium successfully attracted new players over time.

In Article III, we collected 1,431 projective test answers in 28 days (51.1 per day). Even though this task could be categorised as less “worthy” when compared to the one presented in Article I, appropriate task design enabled us to still collect sufficient contributions from the workers while still generating reliable results in an identical deployment environment.

Finally, our situated crowdsourcing market *Bazaar* presented in Article IV collected 62,602 approved tasks in 21 days (2,981 per day). This is considerably higher than all other deployments, possibly due to four factors. First, the input mechanism (tablet instead of public display) is more usable and less prone to fatigue. Second, this technology is more private than a public display meaning that workers felt more comfortable approaching and interacting with it. Third, 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 (Gupta *et al.*, 2012; Heimerl *et al.*, 2012) and quickly lead to workers losing interest. Fourth, the most crucial factor was that this deployment was the only one that offered tangible rewards to workers. This meant that it was able to attract a larger number of workers while also making it more likely for them to come back to complete more tasks. Furthermore, we were able to control labour supply through price-setting as workers exhibited rational behaviour and *Bazaar* worked effectively as a market. Through reward multipliers, we were able to improve task uptake in certain locations, at certain times and complete certain tasks more or less depending on their popularity. Finally, our assessment of workers' use of *Bazaar* involved a situated survey which included a 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 (Sauro, 2014). 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. Many of the survey respondents also 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*.

3.2.2 Quality of crowdsourced work

As mentioned previously, the deployment presented in Article I was the only one to manipulate motivational approaches, and directly compare the performance of situated and online crowdsourcing using the same task. First of all, there was a significant difference in accuracy between workers that were subjected to our motivational approaches. Workers in the control condition (i.e. no motivational approach) were 60% accurate, while workers in the other conditions were over 75% accurate. Furthermore, the fact-checking question effectively identified poor quality work: those that answered it correctly had higher accuracy than those that answered incorrectly (84% vs. 29%). If we were to only consider the most successful motivational approach and those answers that were preceded by a correct response to the fact-checking question, then those workers achieved an average of 88% accuracy. Overall, the accuracy in our deployment ranged between 60-88%, whereas on Mechanical Turk (Rogstadius *et al.*, 2011) it ranged between 66-89%. Even disregarding the fact that in Rogstadius *et al.*'s (Rogstadius *et al.*, 2011) work the majority of workers were paid (3 or 10 euro cents), our deployment produced comparable levels of quality. These encouraging results highlight the fact that it is possible for situated crowdsourcing to produce high quality work.

To investigate accuracy in Article II, we looked at the relevance, as voted by the workers, of the worker-generated keywords in comparison to the keywords generated by three bootstrapping methods: generated algorithmically through LAKE analysis (Kostakos *et al.*, 2013a), obtained by interviewing local citizens, and chosen at random. As expected, this analysis showed that the randomly generated keywords were voted as the least relevant, suggesting that workers for the most part played the game in a serious manner. Results from the automated algorithm (LAKE) highlighted how this approach will only work in certain locations producing substantially worse results in certain heterogeneous places (e.g. library). Unsurprisingly, the methods that leveraged people's local knowledge produced by far the best results, the volunteer and crowdsourced keywords. By analysing the same number of keywords of both methods (top-10 more relevant), we show that the crowdsourced keywords came on top in three of the five locations. In fact, the top-10 crowdsourced keywords in these locations obtained a perfect relevance score (i.e. all votes classified them as relevant). While volunteer keywords can potentially produce better keywords in certain cases, this method does not scale very well requiring substantially more effort in the long run, particularly, if the context in a location changes. Furthermore, we analysed post-hoc the extent to which players agreed with each other in their answers (Callison-Burch, 2009; Ipeirotis *et al.*, 2010). On average, player agreement was 84.2% which highlights the overall consensus when voting on the keywords. This analysis also enabled us to rank the different keywords for each location. A look at worker agreement can be particularly effective to rank contributions for tasks that do not have a single correct answer. In summary, the combination of careful design considerations, crowd-moderation and gamification in our system provided us with an ability to identify undesired input, provide words of high semantic relevance, as well as a broader scope of keywords than other methods.

In Article III, the task was purposely designed to be ambiguous as we were mostly interested in the mistakes done by workers and therefore we were not concerned about task accuracy. However, we still investigated if the technology and deployment environment would adversely affect our findings due to extensive non-serious use. We found that the overall rating accuracy in our field deployment was 56% (compared to 56% in the lab study). These results show that while a deployment in a public setting may be prone to non-serious responses from participants who are "playful" (Hosio *et al.*, 2014a; 2012), in fact this was not the case because accuracy was on par with the lab study. In addition, we also compared the community *emotion index* obtained from our situated crowdsourcing approach to results obtained independently using Day Reconstructions Method (DRM). While results from the DRM were skewed towards the positive spectrum of emotion, mainly due to the fact that participants were aware of the measurement, our analysis demonstrated good correlation between the two methods (R^2=0.68). This result suggests that our approach can accurately capture aggregate emotion of a community.

Finally, for Article IV, all tasks were subjected to two-stage moderation. In the first stage, we rejected obviously flawed responses and in the second stage we relied on the crowd to moderate the remaining responses. During the first stage, we discarded 12,627 of the 75,229 completed tasks (16.8%). This initial moderation and rejection of bad quality work subsequently curbed abuse, and in fact we did notice abusive workers eventually produced high quality work. Also, in a more realistic situated crowdsourcing market scenario, different task requesters would manage their own tasks leading to more attentiveness, and therefore less gaming behaviour. In the second stage, workers completed the moderation task 23,986 times classifying 94% as of good quality.

In conclusion, previous work has shown that not paying workers, as is the case in Articles I–III, can lead to fast but low-quality work (Mao *et al.*, 2013). However, we demonstrate that it is possible to have both fast and good quality work in situated crowdsourcing given appropriate task design and careful considerations of the deployment context, even when no payment is involved. Furthermore, while payment in Article IV could have led to significant gaming behaviour and misuse of the platform, we demonstrate in this work that a situated crowdsourcing market can produce acceptable levels of quality of work.

3.2.3 Summary

Deployments using situated technologies are susceptible to "noisy" results due to unpredictable behaviour from users (Hosio *et al.*, 2014a; Schroeter *et al.*, 2012). Hence, performance in situated crowdsourcing can come into question. However, thanks to our systematic assessment of performance in situated crowdsourcing, we demonstrate that it is generally more attractive than other means of crowd work while providing comparable levels of quality of work, as shown in Table 2.

Table 2. Comparison of performance between the deployments presented in this thesis and deployments reported in literature using different types of crowdsourcing.

Task description	Stimulus	Worker Input	Crowdsourcing Type	Input Technology	Accuracy (%)	Uptake (tasks/day)	Workers
Article I	Image	Text (numbers)	Situated	Public display	60-88	48	n/a
Article II	Current context	Text (short)	Situated	Multipurpose public display	80-90	200	n/a
Article III	Video	Multiple buttons (3)	Situated	Public display	56	51	n/a
Article IV	Image, video, context, text	Multiple	Situated	Public tablet	94	2981	194
Grade exams (Heimerl *et al.*, 2012)	Image	Slider	Situated	Public touchscreen	80	1034	328
Count cells (Rogstadius *et al.*, 2011)	Image	Text (numbers)	Online (MTurk)	Personal computer	66-89	2.2-20	158
Digitilise text (Gupta *et al.*, 2012)	Image	Text (numbers)	Mobile	Personal phone	76-93	1350-2570	221
Digitilise text (Narula *et al.*, 2011)	Image	Text	Mobile	Personal phone	89	500	10
Slashdot moderation (Lampe *et al.*, 2014)	Text	Text	Online (Slashdot)	Personal computer	80	15665	24069
Named entity extraction (Finin *et al.*, 2010)	Text	Text	Online (MTurk)	Personal computer	91	800	42
Translation (Eagle, 2009)	Text	Text and buttons	Mobile	Personal phone	75	n/a	n/a
Reading task (Kittur *et al.*, 2008)	Text	Text	Online (MTurk)	Personal computer	51	210	58

3.3 Emergent worker behaviour in situated crowdsourcing

There is a lack of deliberation on the influence of social dynamics and emergent worker behaviour in situated crowdsourcing. Pragmatically, exploration of these factors can significantly contribute to the improvement of how situated crowdsourcing applications are designed and analysed. Next, we discuss our approaches to tackle this very issue. We highlight the potential for unobtrusive observation of workers in situated crowdsourcing and identify six different behaviours that have an influence on performance. We also discuss the benefits of tracking workers in situated crowdsourcing and how this can be used to identify emergent behaviours.

3.3.1 Directly observing situated crowdsourcing behaviour

Understanding the social dynamics around situated technologies is crucial for the improvement of crowdsourcing in these mediums. By understanding the potential emergent worker behaviours in situated crowdsourcing, tasks can be designed to cater for the more productive worker archetypes. In online and mobile crowdsourcing studies where personal devices are used, workers are normally treated as a black box. There are only a few exceptions in the literature (Rzeszotarski & Kittur, 2011; Rzeszotarski & Kittur 2012) in which task requesters can infer worker performance from the way they conduct the tasks. Situated crowdsourcing offers the possibility to directly observe people completing crowdsourcing tasks. In Article I, we did exactly this in one of the deployed displays, and observed participants' attitudes and social context when completing tasks.

We used video recording to capture all interactions with one of the public displays during the study (Figure 7) which included 123 instances of interaction. Our analysis confirmed instances of the behaviours that we initially noted in our in-situ observations, but also revealed several new behaviours that people exhibited when using the display. The six identified behaviours were:

- *ignorer*: passers-by that ignored the display, exhibiting what is often referred as display blindness (Müller *et al.*, 2009), and
- *unlocker*: those that actually unlocked the screen but completed no tasks. These account for the high number of curiosity clicks mentioned previously.
- *herder*: individuals would approach the display with a group of people, complete some tasks and then leave with the group. The other members would adopt a passive position behind the herder, in a way that suggested they were not applying social pressure but rather observing.
- *loner*: individuals that approached the display alone and typically spent more time than others completing tasks.
- *attractor*: attracted others to join them on the display, commonly referred as the honeypot effect (Brignull & Rogers, 2003), and complete tasks jointly.

– *repeller*: applied social pressure to try to make the worker leave the display. Instances of repellers also happened when groups of two or more people approached the display.

The latter two behaviours ultimately led to a disturbance and delay in the completion of the tasks. In other words, this resulted in the opposite of peer pressure (Kandel & Lazear, 1992) in that workers instead of being pressured to do well, they would engage in performative acts (Schwarz, 2014) resulting in non-serious completion of tasks. Previous work has reported that in some cases the engagement with these interactive public artefacts emerges only when the overall social context provides a "license to play" (Akpan *et al.*, 2013; Jurmu *et al.*, 2014). In the case of playful applications or games, this does not matter and can even act as a catalyst to use (Kuikkaniemi *et al.*, 2011), but when collecting meaningful data from the public, it may be beneficial to attract more loners than groups.

Fig. 7. The different types of behaviour frequently observed around the display. Ignorer: This is the most typical scenario where a passer-by completely ignores the display. Attractor: Person "A" starts using the screen, person "B" becomes attracted and approaches, and eventually "B" leaves, while "A" remains on the display. Herder: "A" approaches and uses the display while a group observes him. Loner: "A" approaches and uses the screen for a relatively long period of time, while passers-by ignore him. Repeller: "A" starts using the screen while "B" uses body language to apply social pressure to "A" to leave. Unlocker: "A" briefly interacts with the display without stopping his walk.

3.3.2 Tracking situated crowdsourcing workers

In-the-wild deployments that use situated technologies typically do not track specific individuals and therefore cannot identify their participants (Kukka *et al.*, 2013). While in many cases this is not an issue, with situated crowdsourcing, identifying workers may be required. The deployments reported in Articles I, II and III did not track workers, even though we conducted opportunistic situated interviews for Articles I and II. Some of the limitations that derive from this include not knowing if a particular worker came back at any point, if certain demographics performed better or not being able to identify and punish abusive workers.

In Article IV, we had the ability to track and identify individual workers due to the registration/login mechanism in place, as well as the follow-up interviews conducted when workers came to collect their rewards. This allowed us to follow individual workers' use of the system longitudinally. Therefore, we were able to know the amount of accounts created daily, the amount of daily logins and if workers returned at a later time to complete more tasks. We were also able to identify workers that reacted to our attempts to control labour supply by changing kiosks, coming at specific times or changing the task they performed, and later confirmed these behaviours in our interviews. Finally, we could pinpoint which workers were abusing the system and issue warnings that they would not be granted HexaCoins for non-serious work, which significantly curbed gaming behaviour.

In terms of emergent behaviours between workers, our interviewees claimed that they used *Bazaar* practically always alone. They reported that it was against their best interests to perform tasks in groups, given that only one would be rewarded. However, friends still formed virtual groups in which they worked towards a joint goal. For instance, the interviews revealed that a group of four workers decided to earn two movie tickets each to go see movies together. They informed each other through SMS and Facebook to take advantage of reward multipliers and achieve their goal as quickly as possible. On the other hand, HexaCoin transfers did not happen often. Based on our interviews, the majority of the 25 recorded HexaCoin transfers were motivated by curiosity, or workers quitting and transferring their HexaCoins to friends. Similar results were obtained in an experimental virtual economy that allowed user-to-user transfers (Takayama *et al.*, 2009). In contrast, user-to-user transfers have been extremely popular in some commercial virtual currency systems, and resulted in significant emergent behaviours (Lehdonvirta & Castronova, 2014). These effects remain to be captured and exploited in a research setting.

Finally, tracking of workers allowed us to implement some game elements to *Bazaar*. For instance, the leaderboard would have not been possible otherwise. In our interviews, workers reported that the leaderboard raised their competitive spirit and motivated them to complete more tasks.

4 Discussion

In this chapter, we revisit the research questions put forth in the beginning of the thesis, and highlight how they were answered. We then discuss the benefits and drawbacks of situated crowdsourcing, and the differences between using non-personal and personal devices for this purpose. We conclude this chapter with some of the limitations of the deployments reported in the articles present in this thesis.

4.1 Revisiting the research questions

4.1.1 RQ1 – What considerations should be taken into account when designing crowdsourcing applications for situated technologies?

Today, poor quality crowd work can often be attributed to poorly designed tasks (Kittur *et al.*, 2013). Designing effective crowdsourcing tasks can be particularly challenging when using situated technologies in public spaces. Previous work has shown that users are prone to misappropriate these technologies, using them mostly in a non-serious manner (Hosio *et al.*, 2014a; Hosio *et al.*, 2012). This is of particular importance for unpaid situated crowdsourcing deployments. It is more likely that workers will endure non-optimally designed tasks when they know they are getting paid for their efforts. This thesis highlights several key points to consider when designing crowdsourcing tasks for situated technologies. First, motivation plays a significant role in the design of situated crowdsourcing tasks. In Article I, we observed a significant difference in performance between workers that were presented with motivational text versus those that were not. Similarly, in Article II, gamification was shown as an effective motivator to attract and engage workers to conduct situated crowdsourcing tasks. Second, we demonstrate in Article II that workers in a situated crowdsourcing environment can effectively evaluate each other’s input and providing a ranking of contributions when the task does not have a single correct answer. Third, we demonstrate that situated crowdsourcing can significantly benefit from a simpler and more engaging task design. This was apparent in Article III where there were much less curiosity clicks and more tasks completed per session when compared to Article I. Finally, in Article IV, we demonstrate that workers in a situated crowdsourcing market exhibit price elasticity, making it feasible to control and adjust labour through a price mechanism. Crucial to this point is that price-setting should follow contextual and cultural factors rather than online trends.

4.1.2 RQ2 – What are the differences between situated crowdsourcing and other means of collecting crowd contributions in terms of tasks uptake and quality of work?

The affordances of situated technologies and appropriate task design enabled a greater task uptake in our experiments than what has been reported in online and mobile crowdsourcing. At the same time, the contributions were of comparable quality to online/mobile crowdsourcing and other means of collecting crowd contributions even in our non-paid deployments. This was observed in Article I in which our situated deployment outperformed the online deployment. In Article II, our situated crowdsourcing approach produced more accurate and richer results than other methods of collecting crowd contributions. Furthermore, our aggregated crowdsourcing results from Article III successfully captured the community's diurnal rhythms of emotion consistent with an independent DRM study, literature on affect and our understanding of the community's daily routine. Finally, our situated crowdsourcing market in Article IV attracted a populous workforce without a noticeable diminishing of the quality of work. In summary, one should not forego these technologies for crowdsourcing purposes assuming that the tasks are not haphazardly designed.

4.1.3 RQ3 – What are the emergent behaviours that crowd workers exhibit when using situated technologies?

A key characteristic of the deployments presented in this thesis and situated crowdsourcing in general, is that the tasks are completed in a public space. Hence, social dynamics are an important element that can influence task performance. In Article I, we had a unique opportunity, for the first time, to systematically observe workers performing crowdsourcing tasks using a situated technology. The findings from these observations highlight how different groups of workers will approach situated crowdsourcing tasks in a different manner. Furthermore, this thesis discusses the benefits of tracking workers of situated crowdsourcing longitudinally. Typically, deployments using situated technologies not to track individuals, but for crowdsourcing purposes, this may be a requirement. We discuss the insights that were made possible from tracking workers in Article IV and highlight when it can be an important feature. We also acknowledge that a mechanism to track workers can ultimately work as a barrier for participation and therefore recommend its use to be limited only to those cases when it is an actual necessity (e.g. when payment is required, studying individual worker performance over time).

4.2 Benefits and drawbacks of situated crowdsourcing

Situated crowdsourcing offers a complementary – not replacement – means of enabling crowd work. The type of crowdsourcing to be chosen will ultimately depend on several different factors. For instance, situated crowdsourcing can minimise the barriers to contribution from a workers' perspective by minimising the initial effort, which leads to a rapid increase in the number of contributions. However, it may require substantial effort from the task requesters, particularly if there is no infrastructure in place, and can incur in higher initial costs when compared to an online server. The majority of situated crowdsourcing deployments will have the technology set-up in an ad hoc manner, which may be a deterrent to collect crowd data in this fashion. In the case of Article II, we were fortunate to leverage existing interactive public display grid in downtown Oulu, while for the remaining articles, we used available technology in our own campus. However, it can be the case that situated crowdsourcing means having an additional key stakeholder beyond the task requester and workers: location managers (Hosio *et al.*, 2014b; Hosio *et al.*, 2014c). These are the individuals who are involved in managing the physical location where situated technology may be or already is installed. Hence, the commitment of location managers is crucial for leveraging their existing infrastructure for situated crowdsourcing or getting permission to deploy such infrastructure. Therefore, the infrastructure or crowdsourcing application must provide value not only to workers, but also to the location managers (Hosio *et al.*, 2014b). A straightforward way would be to give the managers a small percentage of the profits originating from the tasks performed at their location or simply providing entertainment to their patrons. This would also motivate the managers to maintain and advertise the opportunity to their audiences. At the same time, it can offer people new ways to spend time in a mutually beneficial way to task requesters, workers and location managers. An example of a similar exchange in society right now is people collecting cans to get a deposit, which benefits everyone involved.

Furthermore, as we demonstrate with our deployments, a situated crowdsourcing approach should be considered if fast task completion is a requirement. The serendipitous and self-advertising nature of situated technologies (Müller *et al.*, 2010) helps raise the awareness of crowdsourcing among the local community. Even without payment, we demonstrate in Articles I, II and III that situated crowdsourcing is sustained through a self-renewable "workforce" by steadily attracting new workers. Previous work has highlighted how other forms of crowdsourcing, namely mobile crowdsourcing, have difficulties in sustaining participation (Gupta *et al.*, 2012). When the *mClerk* project was initially presented, workers were engaged and participation was high. However, as soon as rewards were reduced, the number of active users dropped 53% within a day. The majority of workers who left reported that they could not invest more time and work for lower compensation, even though they were still getting paid.

Situated crowdsourcing can also be beneficial when specific local or expert knowledge is required. While previous work has shown that it is possible to crowdsource knowledge curation for niche and specialised topics online through an advertising network (Ipeirotis & Gabrilovich, 2014), situated crowdsourcing is more effective for "local" tasks. The theory of *psychological distance* suggests that people are more likely to engage in tasks that are "close" to them psychologically (Liberman *et al.*, 2007) as is the case with locally relevant tasks. Furthermore, the short work durations and small rewards in these online crowdsourcing markets tend to attract specific worker demographics (Ross *et al.*, 2010). Therefore, it is a challenge to recruit workers who speak a particular language, live in a given city (Ipeirotis, 2010) or have domain-specific knowledge (Heimerl *et al.*, 2012). For instance, while the creation of newspaper articles (Alt *et al.*, 2010) and the translation of documents (Zaidan & Callison-Burch, 2011) are appealing crowdsourcing applications, they require workers within a relevant context. In addition, situated crowdsourcing has the potential to reach more "ingrained" knowledge that can be challenging to get online. Mobile crowdsourcing has been used effectively for this purpose by, for example, reaching out to individuals with knowledge on vestigial dialects (Eagle, 2009; Gupta *et al.*, 2012). In Article II, we exemplify this by obtaining relevant keywords to characterise locations where the situated technologies were deployed, something that would be challenging to achieve online.

Another potential benefit of using situated crowdsourcing deployments is that worker contributions can be used to directly study a particular community. Article

III did exactly this by leveraging a situated crowdsourcing task to better understand our university's community collective diurnal emotion. The strong geographical nature of situated technologies enables an effective geographic scoping of the intended community that is being studied. While mobile phones could be used for the same purpose, researchers can be more certain about the context within which the data is collected as direct input situated technologies remain in the same place. Other methods, like social network analysis, can also be used, but they may not provide enough granularity in terms of geographic scope and context of the community being studied.

Finally, as shown in our studies, the complexity of the crowdsourcing task plays a significant role in the success of situated crowdsourcing. For instance, the larger public displays we used in our deployments offer less rich interface controls that can diminish usability when compared to a desktop computer or mobile phone. Hence, as seen in Article I, beyond a certain level of task complexity some workers on these displays will "give up", leading to a decline in performance. This threshold is lower on public displays than on Mechanical Turk, suggesting that harder tasks are more challenging to be reliably crowdsourced on public displays. Other situated technologies with more usable controls, such as the tablets used in Article IV, allow for tasks with more workload (e.g. open-ended survey question or describe the current context) by setting a higher threshold before workers "give up". Furthermore, many of those that do approach these technologies do so with no motive in mind and to pass the time (Müller *et al.*, 2010). Therefore, if the task at hand requires significant workload, then the worker is more likely to abandon the crowdsourcing task all together. However, it is still possible that highly-engaged workers are willing to engage with complex tasks even on bigger public displays. For this to happen, appropriate "scaffolding" needs to be in place, as is the case with some online crowdsourcing projects (Kittur *et al.*, 2011).

In summary, these and other considerations need to be taken into account in order to arrive at an informed decision on what type of crowdsourcing to use. Situated crowdsourcing presents itself as a new and interesting way to collect crowd work that can be beneficial depending on the characteristics of the tasks and goals of the task requesters.

4.3 Crowdsourcing on non-personal devices

A characteristic of situated crowdsourcing that can influence worker performance is that it is typically performed using non-personal devices as opposed to other forms of crowd work. There is a clear distinction between crowdsourcing using one's own personal device (e.g., mobile phone, personal computer) versus a non-personal device that is embedded in a public space (e.g. public displays). A key affordance of performing tasks using your own personal device is that it can be accomplished in the comfort of one's home, enable the use of more familiar technology or even allow the completion of tasks on the go. In most cases, this entails an active pursuit by workers to participate in these crowdsourcing efforts as they typically expect to get paid for their work.

On the other hand, Müller *et al.* argue that situated technologies do not invite people for a single reason, but users come across and start to use them with no clear motive in mind (Müller *et al.*, 2010). Therefore, they reach users that could otherwise be hard or borderline impossible to reach and ultimately have at that moment free time to spare. Workers in our situated crowdsourcing market (Article IV) were completely new to crowdsourcing, and 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. Thus, while crowdsourcing with situated technologies is still just an emerging opportunity, research in the area is encouraging and motivates further exploration. Similar findings have been demonstrated in the past, in the context of bridging citizens and city officials through situated technologies (Hosio *et al.*, 2012). In that study, 67% of the users who used a display to communicate with officials had never before had any kind of contact with them. This further suggests that situated technologies can appeal to a whole new worker base.

Furthermore, we demonstrate that crowdsourcing deployments that leverage situated technologies should be designed with individual workers in mind, because groups of people may exhibit non-serious behaviour when completing crowdsourcing tasks. However, this may be difficult to achieve when using bigger public displays for crowdsourcing as people feel a certain awkwardness and external pressure when interacting on non-personal devices alone in public, where passers-by can observe them using them (Brignull & Rogers, 2003). While previous research on situated technologies argue for designing for group use to mitigate this issue (Hosio *et al.*, 2014a), it can potentially lead to a dip in the quality of crowdsourcing contributions. Co-located crowdsourcing with situated technologies can be rather problematic due to the increased risk of collusion and misappropriation of the crowdsourcing task. The use of tablets in Article IV mitigated this issue to some extent by making the completion of crowdsourcing tasks with situated technologies a more “personal” experience. This is highlighted in our interviews where the majority of workers reported being comfortable performing the tasks publicly due to their body occluding the screen and that they would only approach the kiosks when alone.

Finally, the articles present in this thesis all used non-personal devices. However, this may not always be the case when conducting a situated crowdsourcing deployment. For example, it is possible to simply have a situated technology that works as a pick-up point for crowdsourcing tasks. This way, it is possible to maintain some of the characteristics of situated crowdsourcing (e.g. self-promotion, targeting a specific crowd) while allowing workers to complete the tasks on their own device. Alternatively, this can also be achieved solely through mobile phones by pushing tasks to workers within a specific geographical area. We argue that this approach is more appropriate for more complex or time-consuming tasks that workers would be less likely to want to perform in a direct input situated technology. As with the decision of what type of crowdsourcing to choose, the use of different variations of situated crowdsourcing will ultimately depend on the task and the goals of the task requesters.

4.4 Limitations

The majority of the limitations of the deployments reported in the articles present in this thesis are inherent to field deployments. For instance, in the study reported in Article IV, 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 in our experiments usually lasted just a few minutes.

Further, the walk-up-and-use nature of our deployments resulted in limited usability and accessibility of tasks, with less rich interface controls than a standard desktop environment. This coupled with potential embarrassment or awkwardness when engaging with these technologies in a public place, may have deterred some workers from participating in our deployments.

In the case of the study described in Article I, we highlight that the population we tapped was very different from the population that participated in the compared online deployment. For instance, online workers typically expect to get paid for their work. Work motivated by altruism or to simply to pass the time is much more suitable for situated crowdsourcing than it is for online labour markets such as MTurk. Therefore, we acknowledge the fact that our comparison of the two sets of results in Article I is not ideal, but nevertheless, it is helpful in establishing the overall levels of performance between the different types of crowdsourcing. Furthermore, in crowdsourcing markets, tasks compete for the attention of workers, and therefore, rewards have an important effect on performance. In our case, our deployed technologies did not compete with other in terms of crowdsourcing, but it did compete with all other stimuli in the environment in terms of attention, including the fact that workers were often distracted by other people in the environment.

Further, while we took steps to verify that the technology and context did not interfere with our results (e.g., baseline assessment in Article III), it is still probable that some users performed the tasks in a non-serious manner. However, given our results and, in some cases, their consistency with previous findings, we are confident that these constituted only a small portion of the collected data. In addition, in the deployments reported in Articles I, II and III, we did not track individual workers. This obviously means a lower control in terms of differentiating findings in terms of demographics, getting on-going feedback and following-up with workers through interviews. Finally, novelty can play a significant role when conducting situated crowdsourcing experiments. A situated crowdsourcing approach which deploys situated technologies in an ad hoc manner, like the ones reported in our articles, is more likely to perform better if its deployment time is not overly long. Finally, cultural issues were not investigated, which could affect the acceptability of situated crowdsourcing, particularly in the case of Article IV, given the fact that a reward system was in place.

In summary, due to the aforementioned limitations and others, we further emphasise that our approaches are by no means a silver bullet when it comes to collecting contributions from the crowd, but rather complement existing approaches with a set of unique advantages.

5 Conclusion

This thesis explores the feasibility of conducting crowdsourcing experiments using situated technologies. Throughout the thesis, we present several insights based on empirical results for situated crowdsourcing task design, its performance and the behaviours of its workers. We demonstrate that it is feasible to collect good quality crowdsourcing contributions directly from these technologies, even though much of the previous literature has reported their propensity for non-serious use. We argue that while conducting crowdsourcing deployments using situated technologies can have many challenges, it is still very much worthwhile when task performance can benefit from a situated context.

However, the work presented in this thesis represents only a first step towards providing an extensive and complete assessment of situated crowdsourcing. There are several different research avenues that remain unexplored in this domain. For instance, it is important to explore the use of additional situated technologies for crowdsourcing purposes. The use of QR codes or NFC tags as in-situ task pickup points can provide an additional method of reaching workers to perform crowdsourcing tasks in a situated context while allowing workers to complete the tasks using their own personal devices. Alternatively, self-selected workers could have tasks pushed directly to their devices within a certain geographical location. These variations of situated crowdsourcing enable workers to complete tasks on their own personal devices with all the advantages and disadvantages that come with it.

In summary, we are quite optimistic about the future of situated crowdsourcing. As embedded computing devices become increasingly more and more present in our everyday lives, the effort required for such direct input situated crowdsourcing deployments will greatly diminish, while people's openness to use them will increase. At the same time, there will always be a need from individuals to fill in periods of "cognitive surplus", even when out and about in public spaces. While situated crowdsourcing can be seen as very niche, we argue that it can have an important place as a complementary way of enabling crowd work.

References

99Designs. (2014). Retrieved October 31, 2014, from http://www.99designs.com/

Akpan, I., Marshall, P., Bird, J., & Harrison, D. (2013). Exploring the effects of space and place on engagement with an interactive installation. Proceedings of the 2013 SIGCHI Conference on Human Factors in Computing Systems (CHI '13). New York, NY, USA, ACM: 2213–2222.

Alt, F., Shirazi, A. S., Schmidt, A., Kramer, U., & Nawaz, Z. (2010). Location-based crowdsourcing: Extending crowdsourcing to the real world. Proceedings of the 6th Nordic Conference on Human-Computer Interaction (NordiCHI '10). New York, NY, USA, ACM: 13–22.

Amazon's Mechanical Turk. (2014). Retrieved October 31, 2014, from https://www.mturk.com/

Ananny, M., & Strohecker, C. (2009). TexTales: Creating interactive forums with urban publics. In: Foth, M. (ed), Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City. Hershey, PA: Information Science Reference, IGI Global: 68–86.

Aoki, P. M., Honicky, R. J., Mainwaring, A., Myers, C., Paulos, E., Subramanian, S., & Woodruff, A. (2008). Common sense: Mobile environmental sensing platforms to support community action and citizen science. Research Showcase @ CMU, http://repository.cmu.edu/cgi/viewcontent.cgi?article=1200&context=hcii

Arslan, M. Y., Singh, I., Singh, S., Madhyastha, H. V., Sundaresan, K., & Krishnamurthy, S. V. (2012). Computing while charging: Building a distributed computing infrastructure using smartphones. Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies (CoNEXT '12). New York, NY, USA, ACM: 193–204.

Atkinson, A. P., Tunstall, M. L., & Dittrich, W. H. (2007). Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. Cognition 104(1): 59–72.

Bernstein, M. S., Little, G., Miller, R. C., Hartmann, B., Ackerman, M. S., Karger, D. R., Crowell, D., & Panovich, K. (2010). Soylent: A word processor with a crowd inside. Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology (UIST '10). ACM: 313–322.

Boring, S., Jurmu, M., & Butz, A. (2009). Scroll, tilt or move it: Using mobile phones to continuously control pointers on large public displays. Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7 (OzCHI '09). New York, NY, USA, ACM: 161–168.

Brignull, H., & Rogers, Y. (2003). Enticing people to interact with large public displays in public spaces. Proceedings of the 9th IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT '03). Zurich, Switzerland, IOS Press: 17–24.

Brown, B., Reeves, S., & Sherwood, S. (2011). Into the wild: Challenges and opportunities for field trial methods. Proceedings of the 2011 SIGCHI Conference on Human Factors in Computing Systems (CHI '11). New York, NY, USA, ACM: 1657–1666.

Burke, J. A., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., & Srivastava, M. B. (2006). Participatory sensing. Workshop on World-Sensor-Web: Mobile Device Centric Sensory Networks and Applications (WSW '06), New York, NY, USA, ACM.

Callison-Burch, C. (2009). Fast, cheap, and creative: Evaluating translation quality using amazon's mechanical turk. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP '09). Strodsburg, PA, USA, ACL: 286–295.

Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. Journal of Economic Behavior & Organization 90: 123–133.

Church, K., Cherubini, M., & Oliver, N. (2014). A large-scale study of daily information needs captured in situ. Transactions on Computer-Human Interactions 21(2): 1–46.

Churchill, E. F., Nelson, L., Denoue, L., Helfman, J., & Murphy, P. (2004). Sharing multimedia content with interactive public displays: A case study. Proceedings of the 5th Conference on Designing Interactive Systems (DIS '04). New York, NY, USA, ACM: 7–16.

Clickworker. (2013). Retrieved September 28, 2013, from http://www.clickworker.com

CrowdFlower. (2014). Retrieved October 31, 2014, from https://www.crowdflower.com/

Day, N., Sas, C., Dix, A., Toma, M., Bevan, C., & Clare, D. (2007). Breaking the campus bubble: Informed, engaged, connected. Proceedings of the 21st British HCI Group Annual Conference on People and Computers (HCI '07). Swinton, UK, BCS: 133–136.

Deci, E. (1975). Intrinsic motivation. New York, NY, USA, Plenum Press.

Douceur, J. R. (2002). The sybil attack. Peer-to-peer Systems. Berlin, Heidelberg, Springer-Verlag: 251–260.

Dow, S., Kulkarni, A., Klemmer, S., & Hartmann, B. (2012). Shepherding the crowd yields better work. Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12). New York, NY, USA, ACM: 1013–1022.

Downs, J. S., Holbrook, M. B., Sheng, S., & Cranor, L. F. (2010). Are your participants gaming the system?: Screening mechanical turk workers. Proceedings of the 2010 SIGCHI Conference on Human Factors in Computing Systems (CHI '10). New York, NY, USA, ACM: 2399-2402.

Eagle, N. (2009). Txteagle: Mobile crowdsourcing. Proceedings of the 3rd International Conference on Internationalization, Design and Global Development: Held as Part of HCI International 2009. Berlin, Heidelberg, Springer-Verlag: 447–456.

Etzioni, A. (1971). Modern organizations. Englewood Cliffs, NJ, USA, Prentice-Hall.

Exner Jr, J. E., & Erdberg, P. (2005). The Rorschach, Advanced Interpretation. Hoboken, NJ, USA, John Wiley & Sons.

Finin, T., Murnane, W., Karandikar, A., Keller, N., Martineau, J., & Dredze, M. (2010). Annotating named entities in twitter data with crowdsourcing. Human Language Technologies Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk. Stroudsburg, PA, USA, ACL: 80–88.

Fortin, C., Neustaedter, C., & Hennessy, K. (2014). The appropriation of a digital "speakers" corner: Lessons learned from the deployment of Mégaphone. Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14). New York, NY, USA, ACM: 955–964.

GateGuru. (2015). Retrieved February 17, 2015, from https://www.gateguru.com/

Gibbons, R. (1997). Incentives and careers in organizations. In: Kreps, D. & Wallis, K. (eds), Advances in Economic Theory and Econometrics, Vol. II., Cambridge University Press.

Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. The Quarterly Journal of Economics 115(3): 791–810.

Goncalves, J., Ferreira, D., Hosio, S., Liu, Y., Rogstadius, J., Kukka, H., & Kostakos, V. (2013a). Crowdsourcing on the spot: Altruistic use of public displays, feasibility, performance, and behaviours. Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13). New York, NY, USA, ACM: 753–762.

Goncalves, J., Kostakos, V., Hosio, S., Karapanos, E., & Lyra, O. (2013b). IncluCity: Using contextual cues to raise awareness on environmental accessibility. Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '13). New York, NY, USA, ACM: 17:1–17:8.

Goncalves, J., Hosio, S., Ferreira, D., & Kostakos, V. (2014a). Game of words: Tagging places through crowdsourcing on public displays. Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14). New York, NY, USA, ACM: 705–714.

Goncalves, J., Hosio, S., Liu, Y., & Kostakos, V. (2014b). Eliciting situated feedback: A comparison of paper, web forms and public displays. Displays 35(1): 27–37.

Goncalves, J., Kostakos, V., Karapanos, E., Barreto, M., Camacho, T., Tomasic, A., & Zimmerman, J. (2014c). Citizen motivation on the go: The role of psychological empowerment. Interacting with Computers 26(3): 196–207.

Goncalves, J., Pandab, P., Ferreira, D., Gharamani, M., Zhao, G., & Kostakos, V. (2014d). Projective testing of diurnal collective emotion. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14). New York, NY, USA, ACM: 487–497.

Govindaraj, D., KVM, N., Nandi, A., Narlikar, G., & Poosala, V. (2011). MoneyBee: Towards enabling a ubiquitous, efficient, and easy-to-use mobile crowdsourcing service in the emerging market. Bell Labs Technical Journal 15(4): 79–92.

Greenfield, A., & Shepard, M. (2007). Situated Technologies Pamphlets 1: Urban Computing and Its Discontents. The Architectural League of New York, New York.

Gupta, A., Thies, W., Cutrell, E., & Balakrishnan, R. (2012). mClerk: Enabling mobile crowdsourcing in developing regions. Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems (CHI '12). New York, NY, USA, ACM: 1843–1852.

Harper, F. M., Raban, D., Rafaeli, S., & Konstan, J. A. (2008). Predictors of answer quality in online Q&A sites. Proceedings of the 2008 SIGCHI Conference on Human Factors in Computing Systems (CHI '08). New York, NY, USA, ACM: 865–874.

Heimerl, K., Gawalt, B., Chen, K., Parikh, T., & Hartmann, B. (2012). CommunitySourcing: Engaging local crowds to perform expert work via physical kiosks. Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems (CHI '12). New York, NY, USA, ACM: 1539–1548.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. Experimental Economics 14(3): 399–425.

Hosio, S., Goncalves, J., & Kostakos, V. (2013a). Application discoverability on multipurpose public displays: Popularity comes at a price. Proceedings of the Second International Symposium on Pervasive Displays (PerDis '13). New York, NY, USA, ACM: 31–36.

Hosio, S., Goncalves, J., Kostakos, V., Cheverst, K., & Rogers, Y. (2013b). Human interfaces for civic and urban engagement: HiCUE '13. Adjunct Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13). New York, NY, USA, ACM: 713–720.

Hosio, S., Goncalves, J., Kostakos, V., & Riekki, J. (2014a). Exploring civic engagement on public displays. In: Saeed, S. (ed), User-Centric Technology Design for Nonprofit and Civic Engagements. Springer International Publishing: 91–111.

Hosio, S., Goncalves, J., Kukka, H., Chamberlain, A., & Malizia, A. (2014b). What's in it for me: Exploring the real-world value proposition of pervasive displays. Proceedings of the Third International Symposium on Pervasive Displays (PerDis '14). New York, NY, USA, ACM: 174–179.

Hosio, S., Goncalves, J., Lehdonvirta, V., Ferreira, D., & Kostakos, V. (2014c). Situated crowdsourcing using a market model. Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14). New York, NY, USA, ACM: 55–64.

Hosio, S., Goncalves, J., Kostakos, V., & Riekki, J. (2015). Crowdsourcing public opinion using pervasive technologies: Lessons from real-life experiments in Oulu. Policy & Internet, in press.

Hosio, S., Kostakos, V., Kukka, H., Jurmu, M., Riekki, J., & Ojala, T. (2012). From school food to skate parks in a few clicks: Using public displays to bootstrap civic engagement of the young. Proceedings of the 10th International Conference on Pervasive Computing (Pervasive '12). Berlin, Heidelberg, Springer-Verlag: 425–442.

Howe, J. (2006). The rise of crowdsourcing. Wired Magazine 14(6): 1–4.

Howe, J. (2008). Crowdsourcing: How the power of the crowd is driving the future of business. Random House. Retrieved Octorber 31, 2014, from http://www.bizbriefings.com/Samples/IntInst%20---%20Crowdsourcing.PDF

Hubbard, R. G., & Palia, D. (1995). Executive pay and performance evidence from the US banking industry. Journal of Financial Economics 39(1): 105–130.

Innocentive. (2014). Retrieved October 31, 2014, from http://www.innocentive.com/

Ipeirotis, P. G. (2010). Demographics of Mechanical Turk. Retrieved October 31, 2014, from https://archive.nyu.edu/bitstream/2451/29585/2/CeDER-10-01.pdf

Ipeirotis, P. G., & Gabrilovich, E. (2014). Quizz: Targeted crowdsourcing with a billion (potential) users. Proceedings of the 23rd International Conference on World Wide Web (WWW '14). New York, NY, USA, ACM: 143–154.

Ipeirotis, P. G., Provost, F., & Wang, J. (2010). Quality management on Amazon Mechanical Turk. Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '10). New York, NY, USA, ACM: 64–67.

Jurmu, M., Goncalves, J., Riekki, J., & Ojala, T. (2014). Exploring use and appropriation of a non-Moderated community display. Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia (MUM '14). New York, NY, USA, ACM: 107–115.

Kandel, E., & Lazear, E. P. (1992). Peer pressure and partnerships. Journal of Political Economy 10(4): 801–817.

Kaufmann, N., Schulze, T., & Veit, D. (2011). More than fun and money. Worker motivation in crowdsourcing-a study on Mechanical Turk. Proceedings of the 2011 Americas Conference on Information Systems (AMCIS '11). Atlanta, GA, USA, AIS: 1–11.

Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. Proceedings of the 2008 SIGCHI Conference on Human Factors in Computing Systems (CHI '08). New York, NY, USA, ACM: 453–456.

Kittur, A., Smus, B., Khamkar, S., & Kraut, R. E. (2011). Crowdforge: Crowdsourcing complex work. Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11). New York, NY, USA, ACM: 43–52.

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., & Horton, J. (2013). The future of crowd work. Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13). New York, NY, USA, ACM: 1301–1318.

Kostakos, V., Juntunen, T., Goncalves, J., Hosio, S., & Ojala, T. (2013a). Where am I? Location archetype keyword extraction from urban mobility patterns. PLoS ONE 8(5): e63980.

Kostakos, V., Kukka, H., Goncalves, J., Tselios, N., & Ojala, T. (2013b). Multipurpose public displays: How shortcut menus affect usage. IEEE Computer Graphics and Applications 33(2): 50–57.

Kuikkaniemi, K., Jacucci, G., Turpeinen, M., Hoggan, E., & Müller, J. (2011). From space to stage: How interactive screens will change urban life. Computer 44(6): 40–47.

Kukka, H., Oja, H., Kostakos, V., Goncalves, J., & Ojala, T. (2013). What makes you click: Exploring visual signals to entice interaction on public displays. Proceedings of the 2013 SIGCHI Conference on Human Factors in Computing Systems (CHI '13). New York, NY, USA, ACM: 1699–1708.

Kumar, A., Yadav, K., Dev, S., Vaya, S., & Youngblood, G. M. (2014). Wallah: Design and evaluation of a task-centric mobile-based crowdsourcing platform. Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MOBIQUITOUS '14). Brussels, Belgium, ICST: 188–197.

Lampe, C., Zube, P., Lee, J., Park, C. H., & Johnston, E. (2014). Crowdsourcing civility: A natural experiment examining the effects of distributed moderation in online forums. Government Information Quarterly 31(2): 317–326.

Lazear, E. P. (2000). Performance pay and productivity. American Economic Review 90(5): 1346–1361.

Lehdonvirta, V., & Castronova, E. (2014). Virtual Economies: Design and Analysis. MIT Press.

Lewalter, D. (2003). Cognitive strategies for learning from static and dynamic visuals. Learning and Instruction 13(2): 177–189.

Liberman, N., Trope, Y., & Stephan, E. (2007). Psychological distance. Social Psychology: Handbook of basic principles, New York, NY, USA, The Guilford Press.

Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., & Hartmann, B. (2011). Design lessons from the fastest Q&A site in the west. Proceedings of the 2011 SIGCHI Conference on Human Factors in Computing Systems (CHI '11). New York, NY, USA: 2857–2866.

Mao, A., Kamar, E., Chen, Y., Horvitz, E., Schwamb, M. E., Lintott, C. J., & Smith, A. M. (2013). Volunteering versus work for pay: Incentives and tradeoffs in crowdsourcing. Proceedings of the 1st AAAI Conference on Human Computation and Crowdsourcing. Palo Alto, CA, USA, AAAI: 94–102.

Marshall, P., Cain, R., & Payne, S. R. (2011). Situated crowdsourcing: A pragmatic approach to encouraging participation in healthcare design. Proceedings of the 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '11), Washington DC, USA, IEEE: 555–558.

Mashhadi, A. J., & Capra, L. (2011). Quality control for real-time ubiquitous crowdsourcing. Proceedings of the 2nd International Workshop on Ubiquitous Crowdsouring. New York, NY, USA, ACM: 5–8.

Mason, W., & Watts, D. J. (2009). Financial incentives and the "performance of crowds". Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '09). New York, NY, USA, ACM: 77–85.

Munson, S. A., Rosengren, E., & Resnick, P. (2011). Thanks and tweets: Comparing two public displays. Proceedings of the 2011 Conference on Computer Supported Cooperative Work (CSCW '11). New York, NY, USA, ACM: 331–340.

Müller, J., Alt, F., Michelis, D., & Schmidt, A. (2010). Requirements and design space for interactive public displays. Proceedings of the International Conference on Multimedia (MM '10). New York, NY, USA, ACM: 1285–1294.

Müller, J., Wilmsmann, D., Exeler, J., Buzeck, M., Schmidt, A., Jay, T., & Krüger, A. (2009). Display blindness: The effect of expectations on attention towards digital signage. Proceedings of the 7th International Conference on Pervasive Computing (Pervasive '09). Berlin, Heidelberg, Springer-Verlag: 1–8.

Narula, P., Gutheim, P., Rolnitzky, D., Kulkarni, A., & Hartmann, B. (2011). MobileWorks: A mobile crowdsourcing platform for workers at the bottom of the pyramid. Proceedings of the AAAI Workshop on Human Computation (HCOMP '11). Palo Alto, CA, USA, AAAI: 121–123.

Noronha, J., Hysen, E., Zhang, H., & Gajos, K. Z. (2011). Platemate: Crowdsourcing nutritional analysis from food photographs. Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11). New York, NY, USA, ACM: 1–12.

ODesk. (2014). Retrieved October 31, 2014, from http://www.odesk.com/

Ojala, T., Kostakos, V., Kukka, H., Heikkinen, T., Linden, T., Jurmu, M., Hosio, S., et al. (2012). Multipurpose interactive public displays in the wild: Three years later. Computer 45(5): 42–49.

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. Judgment and Decision Making 5(5): 411–419.

Paulos, E., Honicky, R. J., & Hooker, B. (2008). Citizen science: Enabling participatory urbanism. Foth, M. (ed), Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City. Hershey, PA: Information Science Reference, IGI Global: 414–436.

Prelec, D. (2004). A bayesian truth serum for subjective data. Science 306(5695): 462–466.

Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., & Vukovic, M. (2011). An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM '11). Palo Alto, CA, USA, AAAI: 321–328.

Ross, J., Irani, L., Silberman, M., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers?: Shifting demographics in mechanical turk. Proceedings of the 2010 SIGCHI Conference on Human Factors in Computing Systems (CHI'10 Extended Abstracts). New York, NY, USA, ACM: 2863–2872.

Rzeszotarski, J., & Kittur, A. (2011). Instrumenting the crowd: Using implicit behavioral measures to predict task performance. Proceedings of the 24th annual ACM Symposium on User Interface Software and Technology (UIST '11). New York, NY, USA, ACM: 13–22.

Rzeszotarski, J., & Kittur, A. (2012). CrowdScape: Interactively visualizing user behavior and output. Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12). New York, NY, USA, ACM: 55–62.

Sauro, J. (2014). Retrieved October 31, 2014, from http://www.measuringu.com/sus.php

Schroeter, R., Foth, M., & Satchell, C. (2012). People, content, location: Sweet spotting urban screens for situated engagement. Proceedings of the 2012 ACM Conference on Designing Interactive Systems (DIS '12). New York, NY, USA, ACM: 146–155.

Schwarz, O. (2014). Retrieved October 31, 2014, from http://scholar.harvard.edu/schwarz/publications/subjectual-visibility-and-negotiated-panopticon-visibility-economu-online-digit

Shaw, A. D., Horton, J. J., & Chen, D. L. (2011). Designing incentives for inexpert human raters. Proceedings of the 2011 Conference on Computer Supported Cooperative Work (CSCW '11). New York, NY, USA, ACM: 275–284.

Shirky, C. (2010). Cognitive surplus: How technology makes consumers into collaborators. New York, NY, USA. Penguin.

Stack Overflow. (2014). Retrieved October 31, 2014, from http://www.stackoverflow.com/

Steinberger, F., Foth, M., & Alt, F. (2014). Vote with your feet: Local community polling on urban screens. Proceedings of the 3rd International Symposium on Pervasive Displays (PerDis '14). New York, NY, USA, ACM: 44–49.

Sweetser, P., & Wyeth, P. (2005). GameFlow: A model for evaluating player enjoyment in games. Computers in Entertainment (CIE) 3(3): 3–3.

Takayama, C., Lehdonvirta, V., Shiraishi, M., Washio, Y., Kimura, H., & Nakajima, T. (2009). Ecoisland: A system for persuading users to reduce CO2 emissions. Proceedings of Software Technologies for Future Dependable Distributed Systems. Washington DC, USA, IEEE: 59–63.

Taylor, N., Marshall, J., Blum-Ross, A., Mills, J., Rogers, J., Egglestone, P., Frohlich, D. M., Wright, P., & Olivier, P. (2012). Viewpoint: Empowering communities with situated voting devices. Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems (CHI '12). New York, NY, USA, ACM: 1361–1370.

Topcoder. (2014). Retrieved October 31, 2014, from http://www.topcoder.com/

Ushahidi. (2014). Retrieved October 31, 2014, from http://ushahidi.org

Väätäjä, H., Vainio, T., Sirkkunen, E., & Salo, K. (2011). Crowdsourced news reporting: Supporting news content creation with mobile phones. Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11). New York, NY, USA, ACM: 435–444.

Vlachokyriakos, V., Comber, R., Ladha, K., Taylor, N., Dunphy, P., McCorry, P., & Olivier, P. (2014). PosterVote: Expanding the action repertoire for local political activism. Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14). New York, NY, USA, ACM: 795–804.

von Ahn, L., & Dabbish, L. (2004). Labeling images with a computer game. Proceedings of the 2004 SIGCHI Conference on Human Factors in Computing Systems (CHI '04). New York, NY, USA, ACM: 319–326.

von Ahn, L., & Dabbish, L. (2008). Designing games with a purpose. Communications of the ACM 51(8): 58–67.

von Ahn, L., Maurer, B., McMillen, C., Abraham, D., & Blum, M. (2008). Recaptcha: Human-based character recognition via web security measures. Science 321(5895): 1465–1468.

Waze. (2015). Retrieved February 17, 2015, from https://www.waze.com/

Whittle, J., Simm, W., Ferrario, M. -A., Frankova, K., Garton, L., Woodcock, A., Binner, J., & Ariyatum, A. (2010). VoiceYourView: Collecting real-time feedback on the design of public spaces. Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp '10). New York, NY, USA, ACM: 41–50.

Wikipedia. (2014). Retrieved September 26, 2014, from http://www.wikipedia.org/

Yan, T., Marzilli, M., Holmes, R., Ganesan, D., & Corner, M. (2009). mCrowd: A platform for mobile crowdsourcing. Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys '09). New York, NY, USA, ACM: 347–348.

Zaidan, O. F., & Callison-Burch, C. (2011). Crowdsourcing translation: Professional quality from non-professionals. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Stroudsburg, PA, USA, ACL: 1220–1229.

Original publications

I Goncalves J, Ferreira D, Hosio S, Liu Y, Rogstadius J, Kukka H & Kostakos V (2013) Crowdsourcing on the Spot: Altruistic Use of Public Displays, Feasibility, Performance, and Behaviours. Proc. The 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013), Zurich, Switzerland: 753–762. DOI: 10.1145/2493432.2493481

II Goncalves J, Hosio S, Ferreira D & Kostakos V (2014) Game of Words: Tagging Places through Crowdsourcing on Public Displays. Proc. The 2014 ACM Conference on Designing Interactive Systems (DIS 2014), Vancouver, BC, Canada: 705–714. DOI: 10.1145/2598510.2598514

III Goncalves J, Pandab P, Ferreira D, Ghahramani M, Zhao G & Kostakos V (2014) Projective Testing of Diurnal Collective Emotion. Proc. The 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2014), Seattle, WA, USA: 487–497. DOI: 10.1145/2632048.2636067

IV Hosio S, Goncalves J, Lehdonvirta V, Ferreira D & Kostakos V (2014) Situated Crowdsourcing Using a Market Model. Proc. The 27th ACM User Interface Software and Technology Symposium (UIST 2014), Honolulu, Hawaii, USA: 55–64. DOI: 10.1145/2642918.2647362

Original publications are not included in the electronic version of the dissertation.

518. Jayasinghe, Laddu Keeth Saliya (2015) Analysis on MIMO relaying scenarios in wireless communication systems
519. Partala, Juha (2015) Algebraic methods for cryptographic key exhange
520. Karvonen, Heikki (2015) Energy efficiency improvements for wireless sensor networks by using cross-layer analysis
521. Putaala, Jussi (2015) Reliability and prognostic monitoring methods of electronics interconnections in advanced SMD applications
522. Pirilä, Minna (2015) Adsorption and photocatalysis in water treatment : active, abundant and inexpensive materials and methods
523. Alves, Hirley (2015) On the performance analysis of full-duplex networks
524. Siirtola, Pekka (2015) Recognizing human activities based on wearable inertial measurements : methods and applications
525. Lu, Pen-Shun (2015) Decoding and lossy forwarding based multiple access relaying
526. Suopajärvi, Terhi (2015) Functionalized nanocelluloses in wastewater treatment applications
527. Pekuri, Aki (2015) The role of business models in construction business management
528. Mantere, Matti (2015) Network security monitoring and anomaly detection in industrial control system networks
529. Piri, Esa (2015) Improving heterogeneous wireless networking with cross-layer information services
530. Leppänen, Kimmo (2015) Sample preparation method and synchronized thermography to characterize uniformity of conductive thin films
531. Pouke, Matti (2015) Augmented virtuality : transforming real human activity into virtual environments
532. Leinonen, Mikko (2015) Finite element method and equivalent circuit based design of piezoelectric actuators and energy harvester dynamics
533. Leppäjärvi, Tiina (2015) Pervaporation of alcohol/water mixtures using ultra-thin zeolite membranes : membrane performance and modeling
534. Lin, Jhih-Fong (2015) Multi-dimensional carbonaceous composites for electrode applications

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/

UNIVERSITY OF OULU P.O. Box 8000 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

UNIVERSITY of OULU
OULUN YLIOPISTO

ISBN 978-952-62-0849-7 (Paperback)
ISBN 978-952-62-0850-3 (PDF)
ISSN 0355-3213 (Print)
ISSN 1796-2226 (Online)