



Delft University of Technology

Crowdsourcing

Fast Abundant Flexible User Research for Design

Tidball, Brian

DOI

[10.4233/uuid:c2833bb2-0662-4208-a013-0c084f05f12e](https://doi.org/10.4233/uuid:c2833bb2-0662-4208-a013-0c084f05f12e)

Publication date

2017

Document Version

Final published version

Citation (APA)

Tidball, B. (2017). *Crowdsourcing: Fast Abundant Flexible User Research for Design*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:c2833bb2-0662-4208-a013-0c084f05f12e>

Important note

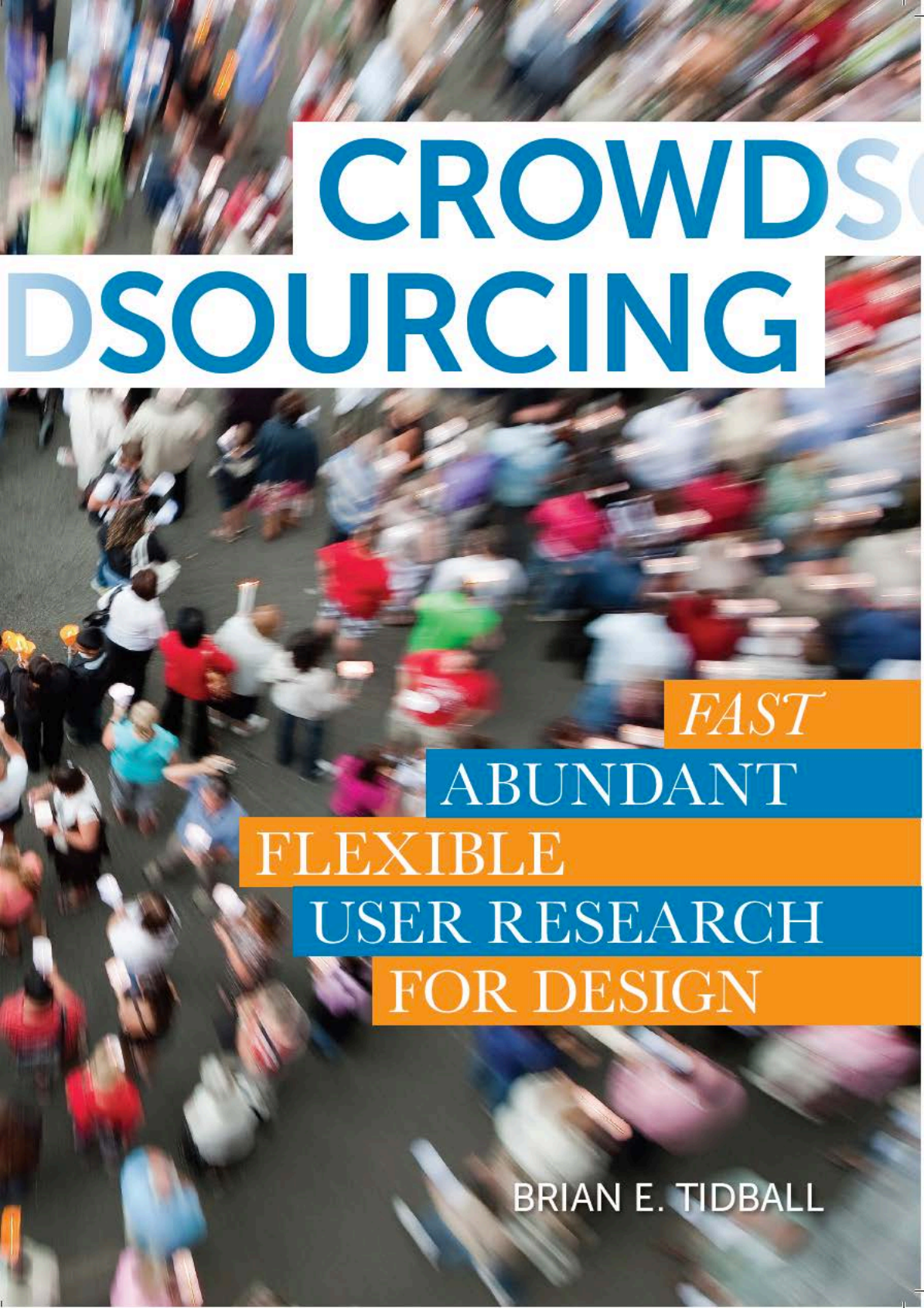
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



CROWDSOURCING

FAST
ABUNDANT
FLEXIBLE
USER RESEARCH
FOR DESIGN

BRIAN E. TIDBALL

CROWDSOURCING:

Fast Abundant Flexible User Research for Design

ISBN ##-#####-#

Keywords: Crowdsourcing, User Research, User Centered Design

* The cover art was crowdsourced on DesignCrowd.com. The design was created by user popdesign, and selected from 33 competing designs.

Layout and design by Corrie van der lelie

© Brian E. Tidball, 2017

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, recording or by any information storage and retrieval system without permission of the author.

CROWDSOURCING:

Fast Abundant Flexible User Research for Design

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben;
voorzitter van het College voor Promoties,
in het openbaar te verdedigen opmaandag 22 mei 2017 om 15:00 uur

door

Brian Esley TIDBALL

Master of Science in Engineering - Wright State University
geboren te Newport Beach, California, De Verenigde Staten van Amerika

Dit proefschrift is goedgekeurd door de promotors:

Prof.dr. P.J. Stappers en dr. I.J. Mulder

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof.dr. P.J. Stappers,	Technische Universiteit Delft, promotor
Dr. I.J. Mulder,	Technische Universiteit Delft, copromotor

Onafhankelijke leden:

Prof.dr. J.P.L. Schoormans,	Technische Universiteit Delft
Prof.ir. D.N. Nas,	Technische Universiteit Delft
Prof. J. Myerson,	Royal College of Art (United Kingdom)
Prof.dr. J. Redström,	Umeå University (Sweden)
Dr. J.V. Khan,	Technische Universiteit Eindhoven



This research was made possible by the financial support of the United States Air Force Institute of Technology (AFIT) Civilian Institutions Program.

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

Preface

40 photos for \$2

When I first experienced crowdsourcing I was amazed by the speed, scale, and diversity of the tasks and activities that people could contribute. I wondered, “Could this be a tool for user research?”

A test of crowdsourcing as an ethnographic user research tool.

After a little more exploring, I decided to test my idea and posted a task on Amazon’s Mechanical Turk (MTurk). In my first attempt, I asked people to submit a picture of what they do to live sustainably. I created a Human Intelligence Task (HIT) on MTurk, and offered \$0.05 USD per photo, for a maximum of 40 photos, and in 3.5 days – with no further effort on my part – I approved their submissions and downloaded the pictures. The result was a mix of personal photos, images from the Internet, and some trash. The collection provided a glimpse into people’s lives and indications of how they perceive the notion of living sustainably. The pictures showed diverse impressions of sustainability that went beyond my expectations; their responses showed a personal relationship between sustainability and their family, employment, and religion.

The forty images submitted by the crowd of workers on Amazon’s Mechanical Turk (a thriving crowdsourcing application) in response to the task: “Please submit a picture of something you do to live sustainably.”



This experience answered my initial question, and demonstrated that crowdsourcing could be used quickly, easily, and affordably to conduct user research. In contrast to traditional user research methods, this on-line crowdsourcing approach changed the relationship between the users and the researcher. In contrast to the trends toward participatory design, crowdsourcing emphasizes anonymity and eliminates steps to recruit or directly interface with participants. The distinct differences save considerable time, and provide an opportunity for designers to reach users in the very beginning of a project. Although crowdsourcing does not allow designers to meet users in person, its speed and reach gives users to a voice early and more often in a user centered design process.

In order to understand the benefits and limitations of crowdsourcing as a new opportunity to conduct user research it is necessary to answer several questions. What changes? What expertise is needed to garner useful responses from users or have participants answer our questions? How do we combine the benefits of user research with crowdsourcing to access and capture the expertise, experience, values, and opinions of people to inform the design process? What are the limitations of this approach?

This dissertation uses an explorative research approach to address these questions and build an understanding of how user research techniques can be combined with crowdsourcing to inform the formative stages of the design process.

Table of Contents

1. Introduction	9
1.1 User Research in Design	10
1.2 Defining Crowdsourcing	14
1.3 Research Focus	16
1.4 Research Approach	18
1.5 Readers Guide	18
2. Understanding Crowdsourcing	21
2.1 Four Types of Crowdsourcing	21
2.2 Research on the Crowd	24
2.3 Ethical Concerns	30
2.4 Building a Framework of Crowdsourcing	32
3. Phase I Studies	37
3.1 Study 1: Crowdsourcing vs. Web search	37
3.2 Study 2: Crowdsourcing Off-Line	50
4. Framing Crowdsourcing for User Research	59
4.1 Reframing Crowdsourcing for User Research	59
4.2 A Conceptual Framework of Crowdsourcing for User Research	60
4.3 Role of the Framework	63
4.4 Research Questions	63
4.5 Supporting Research Approaches	64
4.6 Role of the Researcher	67
5. Phase II Studies	69
5.1 Study 3: Outsource My Job (Logo Contest)	71
5.2 Study 4: Fit In Design (Student Projects)	83
5.3 Study 5: Fast Response (Workshops)	99
5.4 Study 6: Targeting Users (Autism and Connectedness)	108
6. Discussion	121
6.1 Research Implications	121
6.2 Implications for Design	125
6.3 Limitations of Crowdsourcing User Research	128

7. Conclusions and Guidelines	131
7.1 Conclusions for the Framework	134
7.2 Guidelines for Designers	137
7.3 Recent Developments in Crowdsourcing	137
7.4 Issues and Ethical Concerns	140
7.5 Future Research	143
 References	 145
Publications from this Work	151
 About The Author	 153
 Summary	 155
 Samenvatting	 159
 Acknowledgements	 163
 Testimonial	 165

CHAPTER 1 Introduction

In any situation our ability to make informed decisions is dependent on the availability of relevant information. When designing a product, service, or system, an initial understanding of its intended users is essential to ensure their needs are properly integrated into design decisions (Gould & Lewis 1985, Beyer & Holtzblatt 1997). The field of user research provides numerous tools to access various types of information from users to build our understanding and to support design decisions. Unfortunately, the time and expense involved, to engage users often limits user research (Spool & Schroeder 2001). The necessary time and expense can limit or delay the availability of relevant information about users. This delay (or lack) of information can be especially harmful in the formative stages of the design process where early strategic decisions about project direction are formalized (Schrage et al. 1991). In order to bring information about users into the process sooner, fast and flexible methods for conducting user research are welcomed. The continuing evolution of the Internet is providing new opportunities to access users and conduct user research earlier in the design process.

Over the last decade the Internet has created a variety of activities allowing people to create and share information online (Howe 2008). The evolutionary transition from consumption to participation online is commonly referred to as Web 2.0, social computing, or the participatory web. Among the many forms of online participation, crowdsourcing applications are providing constructs to direct and synthesize some of these contributions into a fast, flexible, and affordable means to complete tasks traditionally accomplished by employees (Howe 2006). The speed and flexibility of crowdsourcing provides new opportunities to quickly and easily reach many people who are willing to contribute their expertise to a wide variety of tasks and activities. For user research, crowdsourcing may provide opportunities to seek insights from users to ensure strategic decisions made in the formative stages of the design process are based on relevant insights from users.

The following research will explore the use of crowdsourcing as a vehicle to accelerate user research and bring information to the early decisions in the design process. My personal emphasis on the early stages of the design process stems from my experience working with large-scale

military systems. In these complex systems the biggest issues are often a result of incomplete strategic goals and changing requirements. Both of these issues could benefit from early information from users, to avoid assumptions and clarify user roles and needs early, preventing costly changes later in the process. To this end, the studies that follow, aim to provide designers with the knowledge to access the speed and flexibility of online applications to quickly, easily, and inexpensively conduct user research.

The research starts with a literature review on user research in design and the information needs in the early stages of the design process. The literature review is combined with the first phase of research studies, conducted to distinguish crowdsourcing from other online activities and, more importantly, to unpack the underlying process. The first study contrasts crowdsourcing with web search as sources of images to inform the design process. The second study, examines off-line crowdsourcing and the underlying motivations of participants. The findings are consolidated into a framework, which is constructed to reflect the goals and language of user research in design. The resulting framework serves as a scaffold for the second phase of research studies to investigate crowdsourcing as a user research tool for designers to initiate user research early in the design process. The results highlight the connections in the framework and build our understanding of crowdsourcing as a user research tool. In conclusion, the lessons learned through the studies are incorporated into the framework and presented alongside guidelines for practitioners to use crowdsourcing to access user insights.

1.1 User Research in the Formative Stages of Design

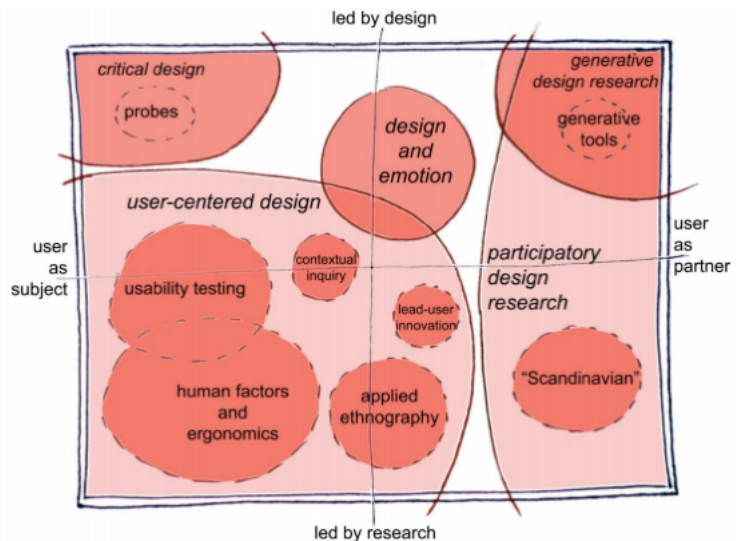
To appreciate the value of information in the formative stages of design we first look at user research in general and then focus on its use in early decision-making. *User research* refers to a wide array of activities that seek out information and increase our understanding of the user in relation to the design of products, services, or systems (Wickens, Lee, Liu & Becker 2004, Kumar 2003, Preece, Rodgers & Sharp 2011). Information about users includes their needs, constraints, preferences, and context of use. It “grounds the [user centered design] process in information about the people who will use the product.” (UPA 2012). The user focus is applied throughout the design process to guide decisions and help ensure the result is a product that is well-suited for the user. In order to maintain a user focus, decisions should to be grounded in information that accounts for the users. Unfortunately, the time and

resources used to conduct user research often limit information about the users, especially early on when project direction is often set.

User research methods elicit insights directly from the end users to inform the design process. This is typically accomplished by studying users and by incorporating them as participants in the design process (Sanders & Stappers 2008). Figure 1.1 depicts the diversity of methods available to engage users in different ways to elicit the desired information. Design literature often refers to these as *design research* methods. In the current work, we deliberately use the term *user research* to avoid confusion with research into the design process. Regardless of the procedure, user involvement provides insights from the users' perspective to inform the design process.

Figure 1.1. Landscape of user research methods being used in the design and development of products and services (Sanders & Stappers 2008).

These user research methods leverage a wide range of tools, including traditional user observation and evaluations common in user centered design process (lower left) to newer methods of early engagement and user participation (right). What is not represented is the time and effort associated with these methods and tools, making use difficult in the formative stages of design.

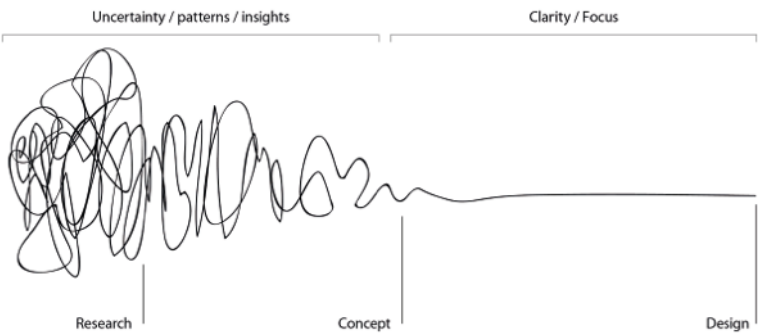


The Formative Stages of the Design Process

The formative stages of the design process are often characterized by uncertainty as the team begins to understand the unique aspects of a new project and research is begun to understand the users, project, and domain. The chaos at the start of the design exploration is characteristically depicted by “that squiggle of the design process” (Newman 2006 - Figure 1.2), where new information is continuously changing the course or emphasis of a project

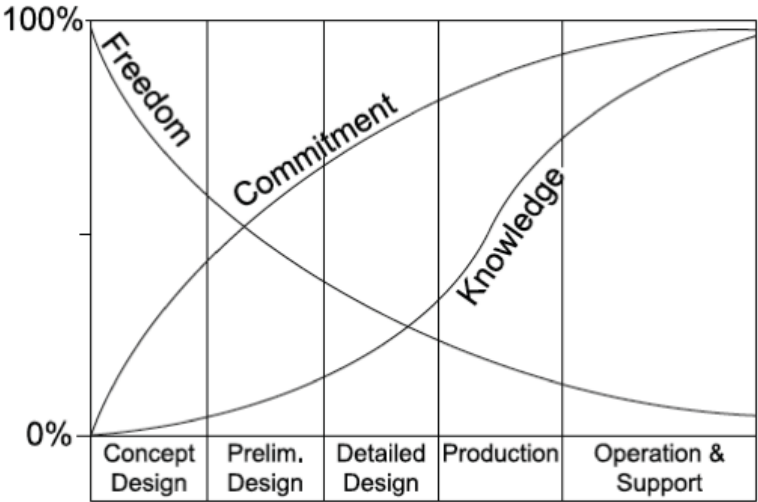
Uncertainty early in the design process is fueled by a combination of a high degree of design freedom and a lack of problem-specific knowledge. Inherent in the formative stages of design is the need to make strategic decisions that move the process forward. Additionally, as a

Figure 1.2. “That squiggle of the design process” depicting the uncertainty in the initial stages of the design process (Newman 2006).



project progresses, time constraints restrict the opportunity for iteration or change. As depicted in Figure 1.3, the necessary project knowledge often lags behind the need to make these decisions (Cooper 2011) and highlights the value of increased knowledge early in the design process. Making the information available prior to key strategic decisions allows for the greatest impact on producing a desirable and successful outcome.

Figure 1.3. The design freedom versus commitment curves, familiar in systems engineering, highlight the knowledge gap early in the design process (Cooper 2011) where problem specific knowledge lags behind the need to commit to decisions. Initial commitments have a lasting impact and are often costly to change, stressing the value to provide user information earlier in the process.



As the design process progresses, the commitment to a design solution and to project investment increases. By increasing our project specific knowledge early on, the knowledge curve moves to the left, ideally ahead of the commitment curve, thus increasing our confidence in design decisions. The current research investigates the feasibility of using crowdsourcing to shift the knowledge curve to the left by bringing initial user insights into the process sooner. The envisioned information from this fast, flexible tool can help guide early strategic decisions and sharpen the focus of later, more expensive user research activities.

Earlier User Information in the Design Process

The user research community has focused considerable effort on accelerating user research. Tools such as rapid software development, paper prototyping, and quick and dirty usability (Nelson 1992) show how initial research efforts focused on simplifying the effort or optimizing the number of participants/users. Limiting participants reduces the workload and cost, which are both barriers early in the design process.

More recently, user research has turned to the Internet to decrease the time and expense of information gathering through automated and asynchronous data collection. Early success has been found through the use of distributed usability testing and digital surveys (Wickens et al. 2004). The adaptation of creative techniques online benefits from online efficiencies and the integration of digital media as stimulus or response, e.g. technology probes (Hutchinson et al. 2003), and online context mapping (Kaptein et al. 2009). These latter examples glean the type of information desired in the formative stage of the design process, though implementation of custom platforms requires additional time and expertise that may not be available early in the design process. While a series of recent books (Kumar 2012, Martin & Hanington 2012, Boeijen et al. 2014) assemble functional method collections, there remains a gap of fast and flexible tools to bring user information into the initial stages of the design process and questions on how – if at all – the low commitment inputs of crowdsourcing could provide the desired insights.

User Information Needs in the Formative Stages of Design

As users and their environments are becoming more intertwined and dynamic in our increasingly connected world, the knowledge needs of designers are shifting to include inspiration, empathy, and understanding (Sanders & Stappers 2008, Hagen & Robertson 2009). Successful design is being achieved by gathering deeper and richer information from users. Researchers seek to engage users in creative activities and increase their role in the design process.

The landscape of methods presented in Figure 1.1 includes several examples of how deeper information is being accessed. Generative techniques engage users in creative activities to unlock rich information (Sanders 2000). Lead User and Participatory Design methods increase the involvement of select users by bringing them into the design process as experts (von Hippel & Katz 2002, Schuler & Namioka 1993). Probes

work to evoke inspirational responses (Gaver, Dunne & Pacenti 1999). Meanwhile Mulder and Kort (2008) exploit emergent technologies for in situ evaluations of user experience in context and leverage the ability to access more users in the daily context of their lives. Tools such as the Personal Card Set (Sleeswijk Visser, van der Lugt & Stappers 2004) and Context Mapping (Sleeswijk Visser, Stappers, van der Lugt & Sanders 2005) help personalize and communicate the rich user information to design teams. While these tools are able to deliver the deep and rich information valued in the formative stages of the design process, they involve substantial time, effort, and expense making it difficult to provide the information in a timely manner.

1.2 Introduction to Crowdsourcing

Among the seemingly endless variety of online activities, applications, and services is a vast array of socially driven, distributed, and participatory activities (Figure 1.4). These activities are often referred to as social computing (Erickson 2013). They provide new and exciting ways for people to create and share knowledge and skills. Crowdsourcing represents one segment that continues to receive attention for its ability to organize and direct large numbers of online participants to accomplish a wide variety of tasks. By incorporating different forms of online participation, inherent in the expanding phenomenon of social computing, crowdsourcing provides a unique structure for directing contributions to produce desired outcomes.

Jeff Howe coined the term “crowdsourcing” in a 2006 Wired Magazine article to distinguish the emerging trend of online production platforms from traditional production models:

“Crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call.” ~ Howe, 2006

The fast, affordable, and flexible characteristics make crowdsourcing appealing as a new tool to gather insights from users to inform design.

Although crowdsourcing is proving to be a powerful method for distributed production and problem solving, there remains confusion about what actually constitutes a crowdsourcing activity (Estelles-Arolas 2012, Shirky 2008). The term crowdsourcing is sometimes incorrectly attached to other online activities where crowds of participants make contribu-

ly choose to participate (or not), submit their *response* or contribution via the *Internet* and receive the predefined *reward*.

Recognizing the desire to bring user insights into the formative stages of the design process, and the potential for crowdsourcing to reach users quickly and easily, the current research will focus on combining the benefits of user research with crowdsourcing to provide fast, abundant, flexible user research for design.

1.3 Research Focus

Problem Statement

Users are a valuable source of information and their insights provide essential contributions to the design process. The ability to access and incorporate user information into the formative phases of a design project helps ensure the conceptual and strategic level decisions are based on an understanding of the users, their needs, and the context of use. Current tools and techniques provide a vast array of options for designers to access and engage users to elicit the desired information. Unfortunately the time-consuming and expensive process of soliciting participants, conducting fieldwork, and interfacing with users, hobbles these techniques. To help ensure user information is readily available, early in the design process, it is desirable for the designers and organizations to reduce the time and expense of conducting user research.

Goal

The goal of the current work is to combine the benefits of user research techniques with existing online crowdsourcing applications to allow designers to quickly, easily, and affordably bring information from users into the formative stages of the design process. This research does not seek to replace or compare crowdsourcing to the valuable insights gained through face-to-face user research methods. Instead, the aim is to augment existing tools with a fast and flexible tool for the early design process to narrow the knowledge gap and focus the design process.

Research Question

In pursuit of the preceding goal, the current research seeks to answer the following question.

How can designers combine the benefits of crowdsourcing with user research to bring information from users into the formative stages of the design process?

This main research question is addressed through a series of six studies, conducted in two phases, as shown in the readers guide (Figure 1.6). Each study was designed to address specific questions, perceptions, and differences to provide the insights necessary for designers to understand and utilize crowdsourcing as a user research tool. Since we are not seeking to replace existing tools, we do not compare the quality of crowdsourcing results to the results of other user research tools. Within the chosen scope, phase I examines crowdsourcing in relation to common approaches of information seeking and examines their benefits and limitations with regard to user research. The two studies in this phase, along with the literature review, provide the insights used to develop a framework and understand the crowdsourcing process. The second research phase includes four studies that examine the value of crowdsourcing to initiate a user centered design in the design process. The research questions in the second phase were largely derived in response to questions and skepticism expressed by designers and design researchers as detailed in Chapter 4, and serve as impetus to build our understanding through exploration within each study.

An exploratory research approach was constructed (as detailed in Chapter 4) to guide the research process. The exploratory research approach was selected to provide opportunities to simultaneously build an understanding of how the process changes and the underlying mechanisms of crowdsourcing as a user research tool.

Scope

The studies examine crowdsourcing as an online approach to provide designers with timely user insights to inform the design process. Specifically it focuses on existing crowdsourcing applications as tools to conduct user research for informing the formative stages of the design process. It seeks a fast, abundant, and flexible means for designers to access information from users. The use of existing platforms leverages the advantages of a readily available and diverse crowd of workers, and system support for creating and hosting the tasks to be performed.

1.4 Research Approach

The studies use a mixed methods approach that combines aspects from three different research methods: Action Research, Grounded Theory, and Design Research. The approach (as detailed in Chapter 4) was assembled to address the research questions, and to build a hands-on understanding of crowdsourcing as a user research tool. Under this

umbrella, the first phase research phase (Chapter 3) works in concert with the literature review (Chapter 2) to provide the insights necessary to develop a framework of crowdsourcing (Chapter 4). The second research phase (Chapter 5) applies a different combination of research methods to address the research questions specific to each study. The tools selected for data collection and analysis are described in the methods section of each study and aim to examine the outcomes from different vantage points. Using a customized approach brings both breadth and depth of information as a means of answering the research questions and building hands-on insight while balancing the limitations or blind spots of a single research method.

1.5 Reader's Guide

The research follows an exploratory process reflected in the chapter outline (Figure 1.6). First, the research begins with a review of literature and practice to build a working knowledge of current crowdsourcing activities. The knowledge gained is used to build a conceptual framework that represents the crowdsourcing process. Phase I then conducts two studies that explore what happens when crowdsourcing is applied to user research. The findings are incorporated into the framework to reflect a designer's viewpoint. The framework is then used as the foundation for the studies in phase II, exploring the value of crowdsourcing for designers. Each study explores different aspects of the employment of crowdsourcing for user research. The first six chapters focus on detailing the research process, while Chapter 7 consolidates the results and presents a set of guidelines for practitioners.

The current chapter defines crowdsourcing and the need to bring user research into the formative stages of the design process. The goal and focus of the research are defined, followed by an introduction to the exploratory research approach.

Chapter 2 presents an overview of crowdsourcing literature, practice, and experience. This knowledge is incorporated into a framework that depicts the current understanding of the crowdsourcing process and contributing factors.

Chapter 3 presents the first of two research phases. The first phase consists of two studies in which we apply our knowledge of crowdsourcing as a work production tool and examine what happens and how crowdsourcing is accomplished for user research. The results make an initial link between familiar research tools and crowdsourcing, while providing insight into what changes when crowdsourcing is adapted for this new context.

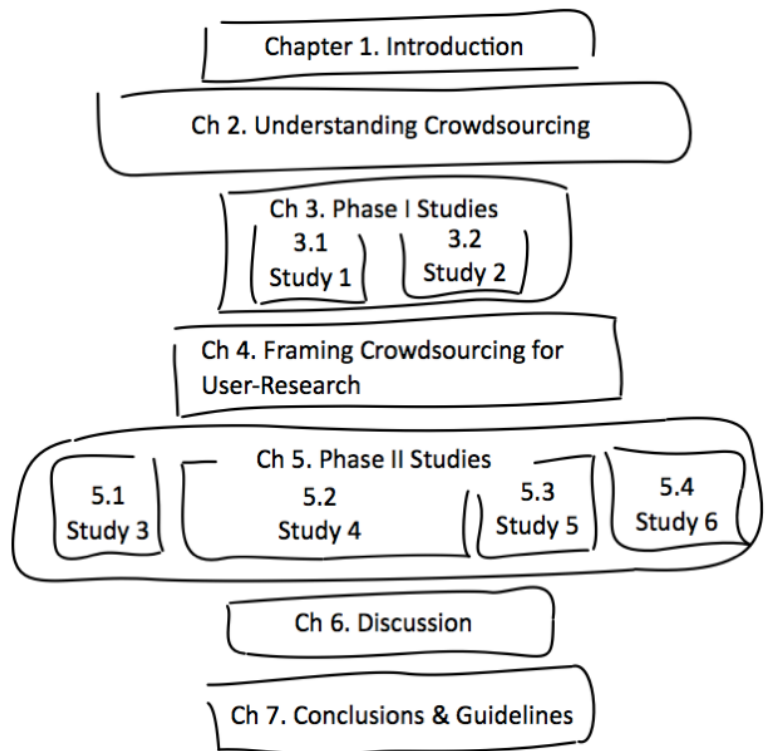


Figure 1.6. Outline of this dissertation

Chapter 4 consolidates the findings from phase I into the initial framework (developed in chapter 2) expanding and adapting it into a framework of crowdsourcing as a means to conduct user research. The conceptual framework serves as the foundation for each study and a means to structure the findings. Chapter 4 concludes with a summary of the supporting research approaches.

Chapter 5 presents the second phase of research studies. Each of the four studies in phase II explores the value of crowdsourcing for designers. The studies in this phase examine different aspects of crowdsourcing as a user research tool. The results further inform the framework and provide insights into how practitioners could incorporate crowdsourcing into the early stages of their design process.

Chapter 6 presents a refined framework based on the research findings from each study. The framework is followed by a discussion of the implications for research and the implications for design. Chapter 6 concludes with a brief discussion of the limitations of crowdsourcing user research as well as reflections on the research process, ethical concerns, and potential for future research.

Chapter 7 consolidates the research findings into a set of guidelines and tips to support practitioners as they apply crowdsourcing to design projects. Chapter 7 concludes with an overview of methodological issues, ethical concerns, and a glimpse into future research opportunities.

Next, Chapter 2 reviews the state-of-the-art in crowdsourcing literature and practice, distinguishing between different forms of crowdsourcing, the demographics and motivations of the crowd, and a brief look at ethical concerns. The insights are consolidated into a framework of crowdsourcing, and defining the elements of the process. Chapter 3 follows with the two studies in phase I.

CHAPTER 2 **A Deeper Understanding of Crowdsourcing**

Before delving into the literature, this chapter examines eight common crowdsourcing applications and organizes them into four categories. These categories provide a view into the underlying process that distinguishes crowdsourcing applications from other online activities. The categories later support the selection of applications to apply as a tool for user research. To date, literature largely focuses on who is in the crowd and what motivates them to participate. In addition, the literature continues to explore the bounds of what is possible. These insights are often applied to optimize the speed and quality of responses from the crowd. Chapter two concludes with the construction of a framework that consolidates and provides a visualization of the crowdsourcing process and sets the foundation for the forthcoming research.

2.1 Distinguishing Four Types of Crowdsourcing

A review and exploration of more than 40 crowdsourcing applications reveals a vast and growing diversity. I found it useful to group the various applications into categories by distinguishing six key elements. While these categorizations do not encapsulate all platforms, they help differentiate among four common types of crowdsourcing: labor markets, problem solving, competitions, and idea markets. Categorization into these four types will also inform the construction of a framework of the crowdsourcing process for user research (Chapter 4).

Table 2.1 (on the following pages) breaks down eight example applications. These eight were selected because they are active and representative of the four primary types of crowdsourcing. The breakdown provides a brief description of six key elements of crowdsourcing, which were identified in Chapter 1 (Application, Requester, Task, Crowd, Response, and Reward). In distinguishing these common types of crowdsourcing, Table 2.1 also provides a glimpse into the diversity in crowdsourcing applications, from many simple tasks completed by many different people difficult tasks presented to a crowd in the hope of receiving one viable response. Unlike Howe (2006, 2008) who makes

distinctions based on who participates, I adopted these categories with a focus on differentiating between what the crowd is being asked to contribute. This distinction focuses on how the process is run and ultimately what the crowd produces, thus providing the designer a sense of how to select and use a platform appropriate for their information needs. While researchers use a variety of terms to categorize crowdsourcing applications, for this research the four types are defined as:

Labor Markets aggregate many small tasks, completed by many members of the crowd, into a single outcome (e.g., translating many sentences and combining them back into a translated book).

Problem Solving focuses on accessing the crowd to find someone who solves a specific (often difficult) problem (e.g., developing a method to eliminate a toxic byproduct from a chemical process, or answering questions that do not have a searchable answer).

Table 2.1 Eight example applications categorized into four common forms of crowdsourcing. The differences are most apparent in the tasks, types of responses, and how they influence the crowdsourcing process

Application	Mechanical Turk mturk.amazon.com	iStockPhoto istockphoto.com	InnoCentive innocentive.com	
Requester	A company or researcher that needs access to a large on-demand workforce.	Designers and other people in need of inexpensive stock images for professional use.	Fortune 500 Companies seeking solutions to technical problems they have not solved internally.	
Task(s)	Large projects (e.g. transcription, photo tagging, questionnaires) are divided into small Human Intelligence Tasks (HITS).	Submit photos and artwork to sell as stock images. No specific tasks assigned.	Find a solution to a difficult technical or engineering problem.	
Crowd	Many individuals with no identified expertise. 40% American, 40% Indian.	Photographers and graphic artists.	Technically competent individuals seeking a challenge or opportunity to share or develop their skills.	
Response	Individually completed tasks that can be aggregated back into a complete solution.	Over 500,000 categorized and searchable images for sale.	A solution to your problem. 50% success rate	
Reward	\$0.01 - 0.05 Higher pay often decreases quality of responses.	Most images sell for \$0.24 - \$24 depending on size. The creator receives 15% - 45% of each sale.	\$10,000 - \$1,000,000 for a proven solution.	
Type of CS	Labor Market	Labor Market	Problem Solving	

Competitions allow an optimal solution to be selected from a variety of responses (e.g., a new organization wants a logo, they offer a prize for the best design, once a number of alternatives are submitted a winner is selected and awarded the prize).

Idea Markets combine the skills of different people through a collaborative process of idea generation and refinement (e.g., a product idea for reducing in-home water usage is submitted, the idea is discussed and refined to make it more appealing to customers and easier to produce).

Differentiating among the different types of crowdsourcing allows the requester to identify applications that leverage the crowd in different ways (be it competition or numerous inputs from a diverse audience) and ultimately select an application capable of producing outcomes that meet specific needs. In general it is the size and complexity of the tasks that drives different reward structures to motivate participation

	Aardvark vark.com	Design Contests hatchwise.com	Crash The Super Bowl Doritos and Pepsi Max Commercial Contest	Threadless threadless.com	Quirky quirky.com
	Subscribed members of Aardvark.	Any company or organization seeking a logo.	Doritos Marketing Department	The Threadless T-shirt company and their community of T-shirt fans.	The Quirky Company.
	Answer questions related to your expertise.	Submit professional quality logo designs.	Submit an "Awesome" commercial for Doritos or Pepsi MAX.	Submit T-shirt designs. Comment and vote on designs. Refine your design based on comments.	Submit design ideas. Comment and vote on other ideas. Refine your idea based on the influence of others.
	Subscribed members of Aardvark who self identify general topics of expertise	People with graphic design skills. Often students and unemployed designers looking to use, develop, or share their expertise.	Creative people with interest in the Super Bowl, Doritos or making commercials. At least basic video production skills.	A community of people passionate about graphic design and T-Shirts.	A community of people with product ideas and "influencers" that want to bring new products into fruition.
	1 or more answers.	A wide variety of logo options to choose from.	Many fully produced commercials to select from. Additional marketing buzz created by the contest.	Refined and vetted ideas with identified customers.	Refined and vetted product designs with identified customers.
	\$0 Helping others in exchange for answers to your own questions.	\$100 - \$500 for a winning design.	Top 10 receive \$25k, Top 6 will be played during the Super Bowl. #1 receives \$1,000,000 #2 receives \$600,000 #3 receives \$400,000	\$2000 + \$500 Threadless gift card for a design that is printed.	30% of direct retail profits + 10% of indirect profits go to the participants. The idea originator usually gets 35% of this, the rest is divided among influencers.
	Problem Solving	Competition	Competition	Idea Market	Idea Market

from the crowd. Understanding these differences also provides a lens to interpret the research literature and later inform the construction of a framework of the crowdsourcing process. Next we look at what literature can teach us about the crowdsourcing process.

2.2 Research on the Crowd

To date research on crowdsourcing has primarily focused on who is in the crowd and what motivates them to participate, often while experimenting with what the crowd can produce. These insights are frequently applied to optimizing speed and quality of responses (Poetz 2012). Literature also examines how participants select tasks (Chilton et al. 2010) and the challenges of organizing crowdsourcing communities (Chi 2010). The growing variety and diversity of the tasks and platforms documented in literature is testament to the flexibility and versatility of crowdsourcing.

Figure 2.1 provides a landscape of the research themes from literature. Following the landscape is a discussion of crowd demographics, motivations, and ethical concerns of crowdsourcing. The demographics of the crowd provide insight into who is likely to respond. Since a person's motivations for responding can have an impact on the quality of their response, understanding motivations will help in the future development of tasks. Chapter 2 concludes with the development of a crowdsourcing framework. The framework provides a view of current crowdsourcing process, and serves as a foundation for Chapter 4 where it is reframed into a tool for conducting user research.

Demographics of the Crowd

In user research, considerable time is often spent recruiting specific participants. In crowdsourcing, a crowd of potential participants is readily available; however, the participants choose which tasks to complete. To make use of this juxtaposition for user research it is first beneficial to understand who is and is not in the crowd. Understanding who is in the crowd will assist the designer in selecting an application, addressing the crowd, and motivating participants to respond. Fortunately, there is a considerable body of literature that has studied who participates in crowdsourcing.

The crowds who contribute are subsets of the “Internet elite” (Brabham 2008). The elite are people who participate online, both as consumers and contributors, in a wide array of online activities. This large and diverse group mimics the demographics of the general population, but with some notable differences. The Internet elite tend to be younger,

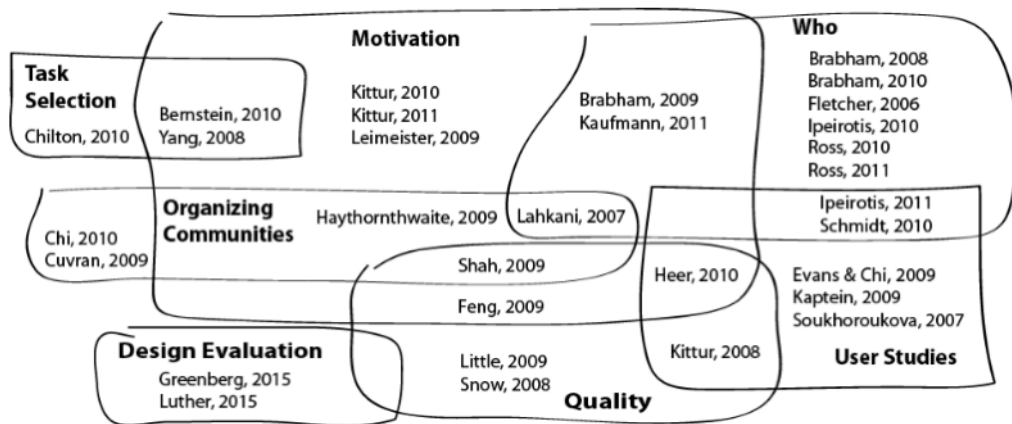


Figure 2.1 A landscape of crowdsourcing literature, grouped into the major research themes. Research has largely focused on who participates, why they participate (motivations), and assessing and refining response quality. These studies cover a wide assortment of tasks and topics across various crowdsourcing applications.

more educated, more international, and highly active online (Brabham 2008, Lakhani & Panetta 2007, Ross et al. 2010). The distribution of household incomes is on par with the general US population (Ipeirotis 2008), though this seems to be less generalizable across platforms. These demographics are evidence that participation is tied to people that have spare time and NOT constrained to those seeking incomes (Ipeirotis 2008 & 2010, Evans & Chi 2009). This decoupling from an income source lends insight into the motivations for participation discussed in the next section. Aside from generalizations across platforms, each application tends to attract a different crowd with unique characteristics. Literature provides demographics for popular applications (MTurk.com, InnoCentive.com, iStock.com, and Threadless.com).

Amazon's *Mechanical Turk* is by far the most studied population, and boasts a diverse registered crowd of more than 500,000 workers in 190 countries (MTurk.com). The crowd is closely aligned to the US population of Internet users, though they do tend to be younger (80% under 35 years old), educated, and female (60%). They also have smaller families, and slightly lower income distribution (Ipeirotis 2008 & 2010, Evans & Chi 2009, Ross et al. 2010, Kaufmann, Schulze & Veit 2011). There is also a diverse mix of professional sectors (education, finance, health, government, etc.) and job roles (manager, assistant, CEO, customer support, etc.) (Evans & Chi 2009). More recently MTurk has seen an increase in Turkers from India, which has steadily grown from 8% to 36%. Indian Turkers tend to be younger (average age 28 versus 35 in the US), male (65-75%), and more educated (70% have a bachelor's degree or greater) (Ipeirotis 2010, Ross et al. 2010). While this large population likely contains the desired knowledge, it may be difficult to target specific members of the crowd.

InnoCentive specializes in hosting difficult problems and offers large prize money for solutions (\$5,000-\$1 million). The crowd on InnoCentive has grown from 80,000 scientists in 150 countries (Lakhani, Jeppesen, Lohse, & Panetta 2008), to 250,000 from over 200 countries in 2011 (innocentive.com 2011). Lakhani, et al. (2008) found that 65.8% of solvers hold a doctorate. The challenges on InnoCentive.com include many specialized technical topics including chemistry, biology, engineering, and manufacturing. Lakhani's research into how and who wins, shows that solvers often have expertise in more than one topic and anecdotally fit the image of tinkerers and inventors. These innovators also demonstrate that the answer may lie outside the expected domain.

iStockPhoto is a popular repository for photographers (who meet community imposed standards of quality) to post their images for sale. Demographics outline the population that is primarily male (65%), married (50%), white/non-Hispanic (88%), North American (54%), European (33%), upper-middle-class (74% indicating "wealthy" or "able to pay bills with extra money to save"), educated (63% have bachelor's degree or greater) (Brabham 2008). In addition contributors are NOT professional photographers. Only 3.9% identify themselves as professionals or designers. Other occupations include self-employed (30%), professional/technical (28%), and student/professor (8%). This profile fits in the category of elite Internet users who participate and contribute online with an observable skill in photography. Sites that require specific skills may provide a more focused crowd; however, they may not provide an ideal structure to generate responses desired for user research.

Threadless is a T-shirt company that relies on a crowd to contribute and select new shirt designs for them to sell. While only minimally studied, this active community is divided into designers (40%) who upload new graphic designs and a larger group of enthusiasts (60%) who engage with designers by commenting on designs, purchasing, and contributing to the blogs. These enthusiasts do not feel they have the necessary skills to express or create their own designs (Brabham 2010, Fletcher 2006). The enthusiast population tends to be young (18 to 30 years old) and evenly distributed between male and female. This division in the crowd is seen in other applications and may help in directing activities to the appropriate subset within an application's crowd.

Crowds contain both committed contributors and transient contributors. Other researchers refer to these groups in different ways and indi-

cate that the underlying motivations are different, but do not indicate how they differ: streakers and samplers (Heer & Bostock 2010), lightweight and heavyweight (Haythornthwaite 2009), short and long term contributors (Huberman et al. 2009). Although there is limited data, long time workers, especially in micro task markets, often subdivide into different groups. Some stick with a specific company (develop trust, familiarity with task type/structure), others stick with a specific type of task (translation, surveys, link checking, etc.), while others participate in a variety or select a few ‘interesting’ tasks.

Considering these examples, we build a sense of how tasks and the design of different applications influence not only what is produced, but also who is attracted to participate and what patterns for participation they follow. With this in mind we can more purposefully select applications that are likely to attract people with the skills and knowledge that benefit our user research endeavors. The size and diversity of the various crowds make it likely that the people or expertise we seek is available.

Motivations of the Crowd

Understanding the motivations of the crowd is essential to understanding why and what people choose to contribute. This insight can help design tasks that generate quality responses from appropriate users. Throughout literature the motivations driving participation in crowdsourcing are coupled to the quality of the responses. Insight into who is participating in the crowd is helpful in understanding the potential knowledge and skills available. More interestingly, knowing why people choose to participate allows us to more effectively communicate our request and entice participation to receive a desirable response. Leimester and Huber (2009), in developing an ideas competition platform, suggest that supporting user motivations is necessary to stimulate both activation and participation in crowdsourcing activities. Knowing why people participate helps us design an attractive task.

Motivations are a mixture of several intrinsic and extrinsic factors, and unique to each individual and situation. Think about all the reasons why people participate in hobbies, games, philanthropy, and employment. Add to this the motivations and desires behind socializing and being part of a community and you uncover motives for joining crowdsourcing activities. Fortunately, there are some underlying trends that can help direct our efforts as we design and host crowdsourcing tasks.

Research on the motivations involved in crowdsourcing fall into three general segments: economic optimization, individuals' motivations on a specific service/application, and non-monetary motivations associated with "open" participation.

Economic Optimization (Speed and quality)

Research on economic optimization seeks to balance the quantifiable elements of monetary rewards and speed against the accuracy or quality of responses. In other words: What is the lowest reward that produces a fast and accurate response? One study examined Task.cn a Chinese "all play auction" site (more commonly known as an online contest) that allows many participants to compete for a single prize. Ignoring all other factors, the results clearly showed that larger rewards generated more participation, both in terms of number of participants and average number of submissions per participant (DiPalantino & Vojnovic 2009).

Other studies seek a "sweet-spot" for paying workers on Amazon's Mechanical Turk. Using an image labeling game, Feng, Besana & Zajac (2009) found that a payment of \$0.05 optimized both speed and quality (agreement with expert labels). Based on the task times this equated to an hourly wage of \$4.39. The results also demonstrated that paying more increased speed, but had a negative effect on quality. While this may seem counter intuitive, higher wages attract spamming and satisficing, meaning that workers will try to collect the big reward without honestly completing the task. This was confirmed by later studies (Kapelner & Chandler 2010). At the optimized reward, the non-expert labels had a 90-96% agreement with expert labels.

A similar study also using Mechanical Turk varied both pay (reward) and difficulty in the classic speed/accuracy task of using a mouse to click back and forth between two changing target objects (Fitts 1954). Using this task, Horton and Chilton (2010) found that less pay produced less work, while longer tasks did not reduce work, indicating that once a participant committed to a task they generally completed it. They also found that on repeatable tasks many workers seemed to work toward target earnings. Using a series of different conditions they calculated a 'reservation wage' of \$1.38/hour. The significant disparity on optimal wages between the two studies raises the question of what other factors affect worker participation and motivation.

Motivations Specific Platforms

A few studies have used surveys and interviews to gain an understanding of worker motivations on different platforms. One of the first studies on crowd motivation, often called the “solvers survey” queried participants on Innocentive.com, a well-known site for difficult research and development problem solving (Lakhai, Jeppesen, Lohse & Panetta 2008). Using an online survey into demographics (as described above), they also surveyed reasons for participation. They found that the three key motivators were: enjoyment solving a tough problem, the desire to win a reward, and having available time. In a related interview, Innocentive.com CEO Dwayne Smith said: “They want to work on problems that matter.” Marc Hodosh (x-prize of Genomics) said: “If you get the public to cheer on scientists like they do runners during a marathon you’ll get this moving.” Hodosh referred to President Obama’s call to increase the use of incentives and contests to stimulate technology innovation (Maxman 2010).

A study of the crowd at iStockPhoto.com showed the most common reported motivations were making money, improving photography skills, seeking a creative outlet, and fun/enjoyment (Brabham 2008). In a separate study, Brabham (2010) found similar results when studying participants at Threadless.com, the user designed T-shirt company. Top motivations were the opportunity to make money, develop creative skills, potential to freelance, and love of the community. A number of participants also stated they had an addiction to the community on Threadless.com, with behaviors similar to behaviors exhibited by people committed to a hobby.

Going beyond monetary rewards, we see analogous motivation structures in areas related to crowdsourcing that lack of monetary rewards. In open development sites like Linux and Wikipedia (Haythorthwaite 2009) and virtual communities like parenting blogs and enterprise social media (Huberman, Romero & Wu 2009, Curran et al. 2009), motivations emphasize recognition, reputation, and intrinsic rewards of skills development/sharing and contributions to meaningful projects.

In her dissertation, Irma Borst (2010) combines several of these ideas on motivation as it relates to both participation and performance. She presents a model of the combined effects of rewards and other motivations on voluntary behavior. The results demonstrate how the lack of monetary rewards can have a strong positive effect on participation and

performance (quality and quantity), assuming you can effectively activate intrinsic and other non-monetary rewards. When people perceive a task to be meaningful and personally rewarding, they are more likely to participate and do a good job. Examples that demonstrate this idea are Linux, wiki projects, Fold It (Khatib 2011), and Galaxy Zoo (Lintott et al. 2008). Each of these examples offers no monetary reward, and derive quality from the participants' desire to contribute their skills to a project they find meaningful or valuable.

These insights into motivation provide a foundation for requesters to use crowdsourcing more effectively. Understanding that it is necessary to balance and incorporate motivations into the task increases the likelihood of reaching workers and producing desired outcomes.

2.3 Ethical Concerns

Ethics and legal concerns are important and widely discussed topics, often mentioned in research and debated in blogs, but remain relatively unstudied in crowdsourcing literature. The biggest concerns are the abuse of workers (inappropriate/illegal tasks, underpayment/"slave labor"), job loss and reduced quality of work (professional photography, graphic design, coding/programming), ensuring regulatory requirements (e.g. Institutional Review Boards, Information Protection Policy) are met to protect private or sensitive information, and unclear intellectual property rights.

From a legal perspective, Intellectual Property (IP) rights are relatively straightforward. When a worker submits a response to a task and then receives the agreed upon compensation, IP transfers to the requester. Simply put, if you pay for it, it is yours. In practice this usually means that if a submission is rejected and not paid, the response cannot be used and ownership remains with the creator. However, IP may be defined differently in the user agreement signed when creating an account for a specific crowdsourcing application. For example, when companies host their own competition, it is common for IP of all content submitted transfers to the company as part of the contest rules. There remain areas that are less defined and vary between applications, such as submissions that are submitted without expectation of compensation (e.g. product reviews). The marketing value of user reviews and contest submissions is significant, and ownership of these ideas is not always clear to participants (Witla 2009). While there is a general standard of paid ownership, details and differences are usually spelled out in the "terms

of use/service” that you must agree to when signing up to participate. In addition to legal ownership of ideas, the intellectual labor of crowds is more valuable than the awards paid to winners (Brahma 2008b). However this free labor is not slave labor as there is no evidence that anyone is forced to participate (Witla 2009). This imbalance between pay and effort can be explained by the diversity of motivations discussed previously, pay is not the only or even primary motivation.

Of specific concern in the research community is the protection of private information, participant and patient confidentiality, and Personally Identifiable Information (PII) in accordance with IRB (Institutional Review Boards) and HIPAA (Health Insurance Portability and Accountability Act) regulations. Most crowdsourcing websites are designed as labor markets and communities of engagement. Therefore, limited attention is given to protecting privacy and many sites intentionally and explicitly identify people and their contributions, for the purpose of giving credit and making connections. Most crowdsourcing applications allow people to use an alias or worker ID and expressly prohibit the collection of PII (birth dates, tax identification numbers).

Illegal activities are a growing concern especially in the micro tasks markets where “click fraud” is occurring. Tasks are posted asking workers to post false product or restaurant reviews or to increase web traffic. Workers have an ethical responsibility to avoid and report illegal tasks; however, it can be difficult to distinguish between legitimate and illegal tasks (e.g. transcribing a CAPTCHA for test or research purposes, or click fraud).

Finally, a lot of online discussion and concern stems from professionals’ fear of losing jobs or clients and diminishing their professions (graphic designers and professional photographers). The new online labor markets now allow non-professionals to participate and at far lower wages than professionals charge. Currently there is no evidence for either lost jobs or reduced quality in the market place. In practice, the opposite appears to be true in the emergence of a new semi-pro, low-cost market for design and photography work.

Each of these ethical concerns is an important consideration and it is the responsibility of the host applications, requesters, and workers to ensure a safe and fair working environment for all. While ethics is not among our research topics, we do maintain an awareness and obligation

to conduct ourselves ethically. Fortunately, the social nature of crowdsourcing applications creates a self-policing effect where the crowd naturally shuns inappropriate actions and hosts encourage participants to report those who breach the terms of service. A reflection on ethics as experienced throughout this endeavor is included in the final chapter.

2.4 Building a Framework of Crowdsourcing

The insights gained through my informal participation and literature review (sections 2.1 and 2.2) informed my understanding of the underlying processes within crowdsourcing. In Figure 2.2 I use this knowledge to expand the initial visualization (left) into a detailed view of the interactions among the requester, platform, and crowd. The resulting framework shows each of the elements in the process and how they interact as a model for production. The goal of the framework is to serve as an initial foundation to explore the use of crowdsourcing as a means to conduct user research.

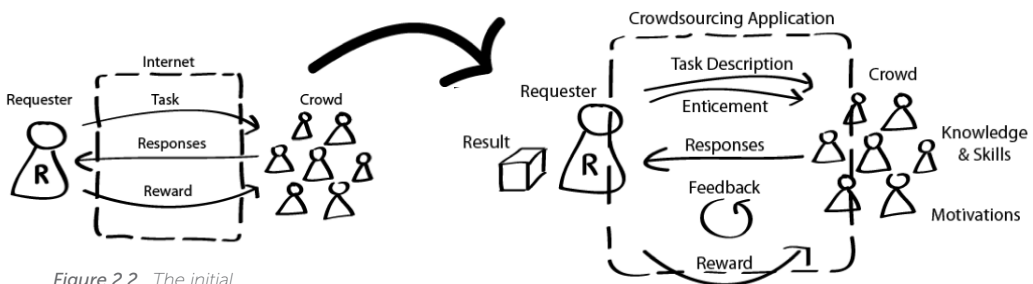
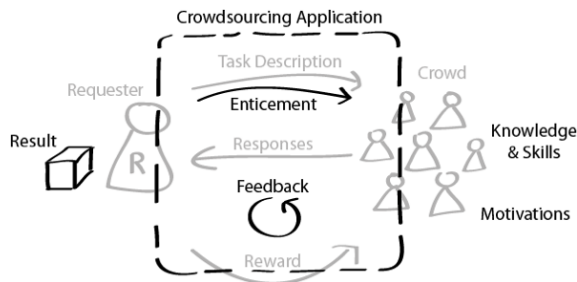


Figure 2.2. The initial visualization (left) of crowdsourcing is expanded into a framework of crowdsourcing as a production model (right). The expanded framework incorporates different elements as identified in literature; calling out the specific knowledge skills and motivations in the crowd, enticement that accompanies the task description, a feedback loop often incorporated into the applications, and a result.

A Crowdsourcing Framework

Figure 2.3 presents a framework of the crowdsourcing process. It incorporates the elements identified from literature regarding: who is in the crowd; their *knowledge, skills* and *motivations*; the effects of different task characteristics (*task description* and *enticements*) on participation; and ultimately producing a desired *result* from the *responses*. In addition we see that the role of the requester goes beyond just outsourcing work. They are able to influence the process in a variety of ways. This new framework provides a comprehensive view of current practice and research, and serves as a foundation for Chapter 4 where it is reframed into a tool for conducting user research.

Figure 2.3. A Framework of the Crowdsourcing Process as a production model. The lighter (grey) elements remain unchanged from the original visualization while the black elements were added to build a holistic representation of current practice. Each element is clearly defined below.



The following list describes each element and its role in the crowdsourcing process (Figure 2.3 above).

1. The **crowdsourcing application** is the heart of any online crowdsourcing process and acts as the intermediary platform between the requester and the crowd. While a wide variety of applications fit the framework above, each is unique. In differentiating their specialties, each application constrains the type, size, and structure of the tasks they support, creating norms and expectations for both the requester and the crowd that participates.
2. The **requester** plays the primary role in the crowdsourcing process. The requester selects the application, defines the task, reviews responses, provides feedback, and distributes the reward(s). Because these responsibilities reside inside the chosen application, a successful outcome is contingent upon understanding and working within the norms and expectations of the application. Fulfilling the role of the requester appropriately is essential to obtaining the desired information.
3. The **task description** specifies what is being requested from members of the crowd and generally includes embedded rewards (intrinsic and extrinsic). The task description is the primary communication from the requester to the crowd and will influence how the crowd responds (quantity, quality, and speed of responses).
4. The **enticement** is an integral part of the task description used to motivate participation. This is where the reward for participation is defined, primarily as monetary rewards. However, the requester should also consider other motivating factors, both explicit and intrinsic.
5. The **crowd** on each application shows discernible characteristics, for any given task only a subset of that crowd will see and then choose to respond. This research is focused on the number of participants, their demographic composition, and their motivations for participation.

Understanding the composition of the crowd helps in selecting a platform and in creating suitable tasks that achieve the desired result.

6. The **knowledge and skills** inherent in any given crowd shift in accordance with the application and the typical tasks requested. Research indicates that the background and expertise is more diverse and deeper than expected. When choosing an application the requester should consider if that application's crowd is likely to include the desired expertise.
7. The **motivations** of participants in crowdsourcing have received a lot of attention. The research has explored why people choose to participate, how they select tasks to contribute to, and the resulting effects on response quality. While the explicit rewards are a motivational factor, research shows that the enjoyment of a challenge and the ability to contribute also play a significant role. This indicates that multiple factors come into play when deciding how to motivate participation.
8. The **responses** are the efforts of individual crowd members to complete the task. Research into responses has primarily focused on increasing the speed and quality. Improvements are seen when balancing rewards with perceived effort, using clear wording in the task description, and the use of quality assurance mechanisms.
9. Most crowdsourcing applications have a **feedback** mechanism built into the platform. Literature into motivations identifies praise and constructive criticism as a form of non-monetary reward, especially for competitions where only one participant gets rewarded. In competitions, feedback is also used to encourage refinement of promising submissions. How-to guidance on many platforms also encourages this from requesters. Feedback is also used to explain why a response will not receive a reward.
10. **Rewards** are paid to responses as defined in the task description/enticement. Research shows that rewards influence the speed and quality of responses, and can be optimized (too small, no attention or responses; too big, spam and reduced quality). Additionally, rewards can include non-monetary compensation such as winning a competition, enjoying a challenge, helping a cause, or receiving gratitude and praise.
11. The **result** is the cumulative outcome of the crowdsourcing event. The individual responses often need to be processed, cleaned, selected,

or aggregated to reach a result to meet the initial need. The final step should be planned and incorporated into the initial task to simplify post-processing and produce a desired result.

In both crowdsourcing practice and literature, prodigious attention is placed on three aspects of the process: the task description (including reward structure), the crowd (demographics and motivations), and the responses (speed and quality).

The task description is the cornerstone of the process. The task needs to be clear, proportional to the reward(s), and account for the norms and expectations of the platform and its crowd. A vague or confusing task will receive limited, inappropriate, or difficult to process results. The same is true for imbalanced or nonconforming tasks. The next chapter presents the first two studies (phase I) that were conducted to build knowledge and experience of the crowdsourcing process while exploring the relationship of crowdsourcing to familiar information seeking activities. Then in Chapter 4, the studies of phase I are combined with insights from the literature review in this chapter to reframe crowdsourcing into a tool for user research. The resulting framework connects current use to user research, highlighting differences in the process and providing the foundation for the remaining research studies.

CHAPTER 3 Phase I - Studies

This first phase of studies seeks to gain hands-on experience with crowdsourcing and develop an understanding of its potential connection to user research. The two studies presented in this chapter explore the use of crowdsourcing in relation to familiar information gathering methods. The research questions for each study were derived to illuminate the underlying process of crowdsourcing, while simultaneously addressing skepticism expressed by designers in early conversation about its fit in the design process.

The first study gathers images from four sources that range from simple web search to crowdsourcing with varying levels of crowd participation. This explores how the process and results of different but related sources compare to one another. The second study uses an off-line crowd to explore the relationship between motivation and responses from a known crowd of participants to the reported motivations of anonymous online crowds. In the next chapter, the insights from these studies are combined with knowledge from literature to reframe the crowdsourcing process into a tool for user research.

3.1 - Study #1: Crowdsourcing vs. Web Search

Designers' skepticism: "Why wouldn't I just search the internet for user information? How is this any different?"

This first study in phase I explores differences between crowdsourcing and web search as sources of user information for design. Increasing participation in social web applications (O'Reilly 2005), including various forms of crowdsourcing, provides a venue for people to contribute to and access a wide variety of user-generated content. These activities open up new opportunities to access user-generated information. By exploring the use of four different online sources for images (MTurk,

Google Images, iStockPhoto, and Flickr) this study seeks to highlight the differences in how designers view the results as well as the process of gathering images for providing user information. This study examines the differences between these sources of online image collection for informing and inspiring the formative

This study is based on the following publication:

Tidball, B., Mulder, I. & Stappers, P.J. (2011). Inspiring Design: Exploring online sources of user-generated information. In *Proceedings of CHI Sparks 2011*, Arnhem, NL, June 2011.

stages of a user centered design process. The results emphasize the distinct characteristics of each source, the designers' preferences, and the value of the crowdsourcing process.

Research Questions

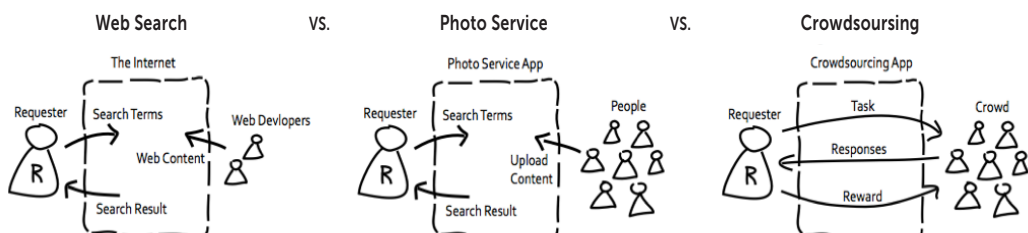
- What distinguishes crowdsourcing from web search as a source of user information for design?
- What sources do designers experience as more appealing to inform design?
- What characteristics make the results from one source more appealing to designers than another source?
- How does the role of the crowd affect the perceived value of the images?

As described in Chapter 2, crowdsourcing is a fast, flexible, and inexpensive method to outsource work. In the context of user research, the speed and economy of an online and hands-off approach provokes questions of how it differs from searching the web. The act of “Googling” has become a natural part of the information seeking process in design; why go to the added effort of crowdsourcing? In Chapter 1 the differences between crowdsourcing and other online activities have been defined. The most notable difference is the participation of the crowd. However, it is similar to web search in that the searcher enters a query and receives results from sources or persons not directly known to them. This similarity obscures underlying differences between the two processes as they relate to conducting user research. Figure 3.1 depicts the differences in participation between web searches, photo sharing services, and crowdsourcing.

Figure 3.1. The differences in how the source influences the results of online user information gathering, this study looks at three sources that primarily differ in why and how the content is provided to the requester: general web content, content on a specific services, and crowdsourcing content provided in response to a task.

Context

This study was conducted as an experiment where pictures were gathered from four online sources. The pictures were gathered on the pre-text of informing and inspiring designers on the topic of “family sustainable living,” a popular topic in design and in society.



The four applications are a source of insight into the everyday lives of people, analogous to cultural probes (Gaver, Dune & Pacenti 1999). The four sources were selected because they accumulate content in different ways, utilizing varying types of crowd involvement.

MTurk uses crowdsourcing to collect inputs from an on-demand workforce in response to a task or request. In this case they were asked to “Please upload a photo of something you do to live sustainably.” Each person received \$0.05 per contribution (Figure 3.2).

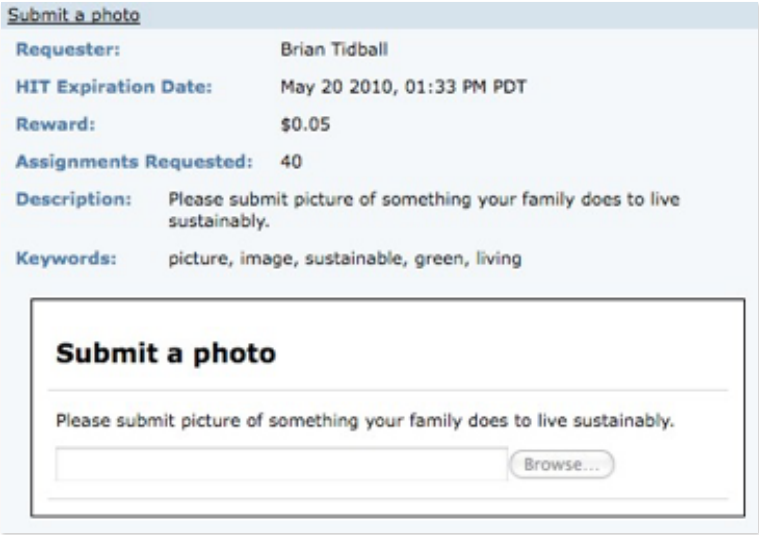


Figure 3.2 Screenshot of the task preview, just before posting the task, showing the set up parameters and task description posted as a HIT (Human Intelligence Task) on MTurk.

Google Images searches out any image on the Internet based on algorithmic relevance to the search terms. In this case the crowd is technically everything that is online, though results will emphasize images from sites that optimize for search-ability.

iStockPhoto is a repository and storefront to buy and sell stock imagery. Primarily amateur photographers populate the site with images and sell the rights to use them. The crowd of photographers is supplying the images with the hope of being paid, but not in direct response to a request. Designers are familiar with iStockPhoto as a relatively inexpensive and readily available source of imagery for graphic and web design.

Flickr is an online community designed to “Share your photos. Watch the world.” (Flickr.com). Images are contributed by people interested in openly sharing and commenting on photographs for non-commercial reasons. A crowd is supplying the images for purely social reasons, with no intent of receiving financial compensation.

Flickr - Interesting: During data collection a second set of images was collected from Flickr using the “interesting” filter. This filter brings images that have received the most attention (comments, likes, shares) to the top of your search query. In this way the crowd has not only supplied the images, but has identified the ones they like best or find most interesting.

Method

To explore the differences and perceived values of these online sources of user information, this study was conducted in two segments. First, image sets were harvested from the four sources, documenting differences in the process and experience. Second, six designers of varying experience were asked to explore and discuss the image sets. These activities were based on the scenario: “Design a product to help families live more sustainably.”

Harvesting the Images

The aim was to gather 40 images from each of the 4 sources: MTurk, Google Images, iStockPhoto, and Flickr x2. An assignment was created on MTurk to gather responses from the crowd while images were collected from the other sources. The assignment was created on MTurk using the standard “file upload template.” The HIT asked people to “Please submit a picture of something your family does to live sustainably.” The HIT offered \$0.05 per upload and allowed only one upload per participant, with no other filters or qualifications. Table 3.1, provides an overview of the setup and parameters used in collecting the image sets along with general data about the responses. In 3.5 days the HIT was completed and the 40 requested pictures were retrieved (see Figure 3.3 in the results).

Images from the other sources (Google Images, iStockPhoto and Flickr) were collected using the default search function on each homepage using the search terms “family sustainable living.” The first 40 images were captured from each source (see Figures 3.3-3.7). While gathering images, Flickr offered three filter options (relevant, recent, and interesting). The default was “relevant” (Figure 3.6), “recent” had little effect on this particular query, while the “interesting” filter produced a different set of results. As this was an explorative study, a second set of images was gathered from Flickr using the “interesting” filter (see Figure 3.7). “Interestingness” is determined by a variety of factors including number of views, comments, shares, and tags a photo receives (Flickr.com).

Table 3.1 Overview of the setup and general results of the image collection activities.

	Crowdsourcing	Search
Application(s):	MTurk	Google, iStock, and Flickr
Title:	Submit a Photo	na
Task:	"Please Submit a picture of something your family does to live sustainably."	Search Terms: "Family Sustainable Living"
Reward/# Requested:	\$0.05 each/40	na / 40
Task Duration Max Work Time	45 min	na
Scheduled Run Time:	5 days	na
Date, Time (CET):	25 May 2010, 14:06	25 May 2010, \pm 14:30
Responses Received:	40	40, 7, 40 **
Responses Rejected:	0 *	\approx 30 million, 0, \approx 3000 ***
Average [Work] Time/Hourly Rate:	4 min/\$0.74	na/na
Completion Date, Time (CET):	27 May 2010, 00:26	25 May 2010, \pm 14:30
Runtime:	34hr 20min	0.26 sec

* Chose to accept all submissions

** Only 7 results available on iStock

*** Additional results were ignored

Throughout the harvesting of the images, specific attention was given to documenting differences in the process and how each experience influenced the perceptions of images from that source. These were strictly researcher observations and they will be noted separately at the end of the discussion section.

Designer Evaluation

To evaluate the different image sets, six designers individually examined each of the five sets in random order. The six designers were equally divided between male and female, all were part of the industrial design program at Delft University of Technology (2 masters student, 3 doctoral candidates, and 1 faculty). All the designers had academic training and professional experience in industrial design; three had specific expertise in sustainable design. They were not informed of where or how the images had been obtained until they had completed the sorting and value judgments for all five sets.

The designers were individually asked to create a new product to support families in living more sustainably. They were then given each of the five sets of images and asked to conduct a card-sorting exercise (Nielsen & Sano 1994) for each set using printed copies. They were given complete freedom in how they chose to sort the images. This allowed the designers to freely explore the images; the act of organizing stimulated them to evaluate each image for meaning and determine

relationships between images (Keller, Stappers & Vroegindeweij 2004). During the card-sorting task the researcher took notes as the designers freely commented on what they saw, how they interpreted individual images, relationships and contrasts between images, and qualities that made individual images more or less interesting or appealing.

After the sorting tasks were completed, the designers were asked to reflect on their preferences and perceived value of the image sets for providing information and inspiration related to the design brief. The researcher's notes captured the designers' comments throughout each session and were later sorted and organized into emerging themes. The comments shown in the results section below (Table 3.2 and 3.3) are quotes that represent the most common comments and reactions to the image sets.

Results

The results for this study include the five sets of images, researcher observations from the harvesting process, and finally designers' perceptions of the image sets as sources of inspiration and user information.

The Image Sets

The first set of images was collected from MTurk. After posting the task it took 3.5 days to receive the 40 requested images (Figure 3.3), at a cost of for \$2.20 (40 responses x \$0.05 + 10% fee). Two of the responses were rejected (not paid) because the file did not include an image; these were automatically reposted and replaced by responses with images. This set has the most color diversity, includes several religious icons, and many pictures of families. The second set of images was from the Google Image search (Figure 3.4). The color green and pictures of architecture dominate the results. The third set of images was from iStockPhoto. The query returned only seven images that matched the search phrase "family sustainable living." The last two sets are from Flickr and include many pictures of people in outdoor settings. The default set includes sub groups of images taken in the same setting, while the "interesting" set has greater diversity.

Each source produced a visibly unique set of images with different characteristics despite use of the same search phrase. There are a few images that appear more than once (Google) or in more than one set (Flickr). The designers' comments below provide insights into the differences they saw and how their observations influenced their perceptions of each source.



Figure 3.3 Images submitted by workers on Mechanical Turk (n=40)



Figure 3.4 Images gathered from Google Images search (n=40)



Figure 3.5 Images gathered from iStockPhoto search (n=7)



Figure 3.6 Images gathered from Flickr search (n=40)



Figure 3.7 Images gathered from Flickr using the 'Interesting' filter (n=40)

The second phase of this study took a closer look at these image sets by presenting them to designers, to explore, assess and describe their impressions of the image sets as a source of information for design.

Assessment of the Image Sets

Throughout the card-sorting task, the designers made many comments as they explored the images. Their comments were relatively consistent and expressed the distinct characteristics of each image set. The comments below are a representative sample of designers' comments for each set.

Table 3.2 Representative Quotes from designers describing the set of images from each source

Source	Pros	Cons
Mturk	"developing vs. the west... reconsider who is consuming" "shows relationship of sustainability to social practice and tradition... a part of life"	"some obvious crap" "lacks 1st world contrast" "lots of people, but not doing things" "too easy to categorize" "no connections [between categories]"
Google Images	"some nice links between groups" "nature, industry, technology" "community projects" "eco can be modern and sexy"	"lacks people" "overview, limited details" "big gaps" "fluff" "I don't want others' solutions" "not homogenic [sic.]" "sustainable is more than architecture"
iStockPhoto	"well composed pictures"	"lack quantity" "how does this relate" "obviously stock photos... not personal"
Flickr	"idealized, but helps make a vision" "interesting contrasts (recycling vs. nature, research vs. life, the good vs. the bad)" "shows all aspects, success and failure" "if only nice pictures you can miss a dimension" "lots of people... doing something is interesting"	"industrial... lacks depth" "they feel forced" "missing tedious photos"
Flickr 'Interesting'	"interesting clusters with good links" "not what first comes to mind" "nice to see people doing stuff" "homogeneous set" "makes a nice infoHgraphic" "not redundant" "more inspirational"	"no city stuff, are we all living rural?" "all happy brown people" "far from subject, but interesting"

Many of their comments described what they liked and disliked about individual images as they related to the topic of sustainable living. In the context of using the images as a source of information for design, their comments show a preference for images that were personal and showed people “doing things” (see Table 3.3). They also appreciated contrast, tension, diversity, and links or relationships among images. These comments hint at the designers’ appreciation for the contextual insights that could be gained from each image set.

Table 3.3 Designers’ comments on what makes a good or poor image or set of images, as a source of inspiration and contextual insights into users.

Positive	"People doing things is nice" "Real snapshots of life" "Doing' something is interesting" "Unexpected" "contrast is nice...western vs. developing cultures" "Interesting'tension...'the'good'and'bad'of[post'industrialization]"
Negative	"Questionable relationship to'sustainability" "I don't want to see typical solutions" "I see people, but they are not doing anything" "stock photos"

The two image sets from Flickr were appreciated over the other sets for their quality, diversity, and depictions of real people and contrasting

ideas. On the other side, the small sample from iStock was correctly perceived as stock imagery and, despite high photographic quality, spurred little to no interest or insights into users. Similarly, Google Images resulted in largely staged interior design and architectural images showing few people. The image set from MTurk received mixed comments. Although not the preferred set, MTurk generated numerous questions and garnered additional attention as the designers worked to interpret and categorize the highly diverse and unexpected images. The added difficulty of sorting the images from MTurk raised many more questions and challenged some of their views of sustainable living. Two of the six designers found the extreme diversity to be a distraction, while the other four commented on how the unexpected images while occasionally frustrating, also aroused deeper thoughts on the topic. This was especially true trying to tie the unexpected religious iconography into their understanding of “family sustainable living.” Separate from their preferences was an empathic engagement with each image seeking insights that resonate and challenge their own knowledge and experience.

Discussion

This study examined four online sources of images as a source of user information. The results demonstrate how the (lack of) involvement of a crowd influences the perceived value of the images as a source of information. Four of the five sets provided both information and points of inspiration for the designers. The set from iStockPhoto provided little of interest, because the images were staged and there were not enough images. To varying degrees, the images in the other sets broadened the designers’ vision of “family sustainable living” and challenged some of their initial assumptions. Their comments highlight image characteristics that affect the perceived value of the photos as a source of information about users. Specifically, the designers preferred photos that were personal and showed people “doing things.” They also preferred sets that showed contrast, diversity, and links or relationships between ideas. They did not appreciate posed or stock photos that showed what they perceived to be someone else’s solution or were not “real” people.

Overall the designers preferred the two image sets from Flickr with a slight edge to the “interesting” set, as they felt the images showed real people doing things. However, it was the set from MTurk, though not as appealing, that triggered the most questions from all six designers. Two of the designers did not appreciate this set, while the others recognized how the incredibly diverse and unexpected images evoked deeper

consideration of their users and their connection to living sustainably. The images from Google were far less personal and as a result were not nearly as informative.

These results show how the different types of crowd involvement influence the output. In the case of Flickr, the crowd strictly posted the images in the interest of social interaction. As a result, the images emphasized real life experiences that provided designers with the contextual insights they found informative. The addition of the “interesting” filter highlighted the images that speak to a bigger audience. In contrast, the images from Mechanical Turk were individually submitted in response to our request. While these images were selected from readily available sources, they were selected specifically to answer the specific task. These distinctions demonstrate that additional value may reside in using multiple sources and considering the differences in the results.

In contrast to online search, the time delay makes MTurk considerably slower (three days versus nearly instant). However, it is much faster and requires less effort than user research methods involving fieldwork or participant solicitation. The image set from MTurk may not produce the richness and personal connection with users that can be gained through fieldwork, but it does offer an affordable option to access first-hand responses from real people without leaving your desk.

After completing the sorting task, the designers were all interested in where the images had come from. This led to discussions about the additional information that is available from each source by looking at the comments or descriptions associated with each image. The fact that the images from MTurk were responses from real people give the image set increased credibility. These insights show that online sources contain additional information that designers can use to gain rich insight into users’ lives. .

Observations from Harvesting

During the harvesting process, as the researcher I noted some differences in how the process influenced my reactions and impressions of the images. The familiar procedure of searching images on Google, iStock and Flickr was no different than any other online search. The final and non-standard step of saving the first 40 images required an additional 10 minutes per set, and while this is relatively negligible it focused attention on each image. In addition, the constraint of saving only the first 40

images went against to my natural reaction of quickly scrolling through the list until something caught my interest.

The additional steps of crowdsourcing to setup, create, fund, and post the task on MTurk altered my perceptions and commitment to the process. While the process takes a little more effort than making an online purchase, the realization that I was asking people to respond rather than perusing images already available created a subtle and unexpected commitment even before I saw any responses. In addition, the time lapse between posting the task and receiving the results, combined with the recognition that a person rather than a computer algorithm selected each image, heightened the attention given to the responses. This was especially experienced for those images that did not match my expectations. Instead of immediately dismissing them as flawed/spammed results, they raised questions of: Why this image? How does this fit? What is the connection? These experiences gave an increased sense of meaning to the photos and sparked an interest in the people who submitted them.

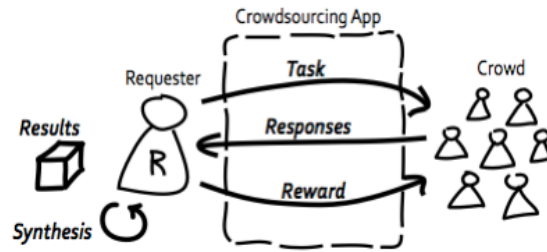
Implications for the Framework

This study shows how the involvement of the crowd takes a deeper look into the information that is available online. For the framework we see that different forms of the crowd's involvement, purpose, and aims of a platform fundamentally alter what is produced and what it conveys to designers. These distinctions are apparent in the designers' preference for sources that are more closely tied to people (Flicker, MTurk).

For the purpose of informing design, Flickr provides a searchable collections of images curated by the crowd for social engagement. And while the images were not intended for our purpose, the underlying social interactions encourage sharing images that are personal and provide contextual information, both of which are beneficial to designers. In contrast, MTurk bridges a gap between the availability of content shared for a different purpose and the time and effort involved in engaging participants. MTurk provides an opportunity to gather inputs online in direct response to your information needs.

It is the structure (though minimal) imposed by crowdsourcing that provides two distinct benefits over online web search as a source of user information. First, is the process of asking people (not searching) for information that encourages the researcher to pause and consciously

Figure 3.8 The value of online information is linked to the reason it has been shared online. For crowdsourcing there is added value in the role of the crowd and the process of asking (not searching) for information, providing a connection to people and their responses. It is also observed that synthesis and post processing on the part of the designer are distinct steps in this information seeking process.



consider what information they are seeking. Second, the realization that you are asking real people to respond to your task, and not simply accessing information posted for different purposes, increases the connection to the results. They are not something found by a search algorithm, but human (and potential end user) responses to a specific request.

Study 1: Crowdsourcing vs. Web Search: Tips for Design Practice

Based on the results and experiences from this study the following tips and lessons learned are shared to help future designers in adopting and making the best use of crowdsourcing in their own design projects.

Tip 1: The formulation of a task for the crowd is an opportunity to consider what information is truly needed.

Tip 2: Take the time to examine all responses from the crowd. Considering why people respond in a certain way may provide insight beyond the response itself.

Tip 3: Take the time to explore the reasons behind search results. Search is so easy it garners little commitment from the designer; quickly scrolling through the results can miss information attached to images in the form of comments and context.

Tip 4: Waiting for responses from people in the crowd reinforces that they are not just search results, and increases the connection to the responses and the information they provide.

Tip 5: Consider the limitations of a platform. There is a threshold where tasks are not appropriate for a selected platform. "Click workers" on MTurk cannot be expected to get up from their computer.

Tip 6: Each crowdsourcing task is a "request" not a "requirement" and should to be treated as such and align with norms on the platform.

Tip 7: When seeking images to inform design, use sources that are social (e.g. Flickr, Instagram) or sources where users are responding to a request (e.g. MTurk).

Tip 8: On MTurk, people will not get up from their desk to take a representative picture. They will select from existing images either from a personal collection or online.

3.2 - Study #2: Crowdsourcing Off-Line

Designers' Skepticism: "Why ask a crowd of online strangers when I have an audience at hand? What differences should I consider?"

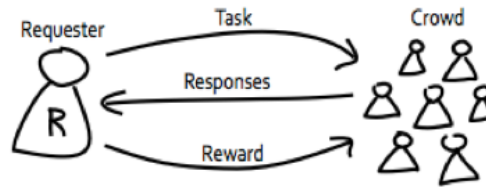
This study examines the use of an existing physical audience to source contextual insights from users. This provides a point of contrast between online crowdsourcing and a more familiar way to access participants. One of the big advantages of online crowdsourcing applications is eliminating the steps to find participants. A convenient method to decrease this burden is to reach out to our social network or a known audience as a source of participants. In this case, a class of 200 industrial design students functioned as the crowd of potential participants. Rather than spamming friends and family with a survey request or seeking live audience participation (e.g. voting), we have defined a task where they can contribute at a later time, allowing the crowd the choice to participate. By examining the processes and motivations governing online participation in crowdsourcing when applied to an off-line crowd (of students), our goal is to understand the role of the platform in crowdsourcing activities.

Research Questions

- What distinguishes online crowdsourcing from questioning a known audience (or off-line crowd)?
- *In the absence of a crowdsourcing application what happens between designer and crowd?*
- *How does moving off-line affect motivations for participation and quality?*

As summarized in chapters 1 and 2, literature has primarily examined the role of motivations in relation to participation and quality of online crowdsourcing applications. While the earliest examples of crowdsourcing (Howe 2006) occurred prior to the advent of the Internet, current literature does not address the differences between motivations and participation on- and off-line. In addition, off-line crowd participation historically emphasizes the *all-play auction* format where the first or best solution wins. In contrast this study examines the aggregate approach where each participant chooses to contribute his or her small piece to the greater whole. While there are observable similarities in the structure of the process, does an offline audience share enough similarity to digital crowdsourcing to operate in a similar fashion?

Figure 3.9 This figure visualizes crowdsourcing with a physical crowd where there is a shared physical connection, but lacks support of an online platform.



Context

This study was conducted in conjunction with an ongoing research project planning future workspaces based on the relationship between a person's digital and physical environments. The lead designer on the workspace project was a doctoral candidate with significant design experience. That project hypothesizes that there are recognizable similarities in organizational structure of workspace elements, such as cluttered versus organized. To test this idea the project wanted to gather image pairs of personal computer desktops and at home workspaces.

Past experience with crowdsourcing on MTurk showed that this platform was a good fit for efficiently gathering these images. However, although the API (application programming interface) on platforms like MTurk would accommodate the task, the effort expected of the participants (retrieving a camera, taking, uploading, and submitting the photos) is far greater than typical tasks on MTurk, which involve a few mouse clicks. In the search for an appropriate platform, we were presented with an opportunity to explore the use of a physical off-line crowd, to gather information from users. We opted to use the largest course available ($N =$ exactly 200 registered students) as our crowd.

Method

The students in one of the largest lectures in our faculty were selected to be the crowd. They were judged to have the tools necessary for participating and there was an acceptable communication structure in place (email). This was necessary to replace the communication functions regularly provided by a crowdsourcing application. After coordinating with the course professor, at the beginning of a lecture we asked the students to support our study. We described the task of taking and sending two pictures, and a flyer was handed out as a reminder of the requested task, (see Figure 3.10) and orally described. The students were asked to email two images: a picture of their computer desktop, and a picture of their workspace. No physical or monetary reward was offered, the only reciprocity was the offer to share the results of the study in a later lesson. The professor reiterated that participation was appreciated, but would not be graded.

Figure 3.10 A copy of the physical task description given to students as a printed postcard requesting their participation and providing instructions.



Those students who emailed their pictures received a follow-up email thanking them for their contribution and asking a single follow-up question “Why did you choose to participate?” The results were compiled and are presented in the results section below.

Table 3.3 A summary of the details used to setup the crowdsourcing activity and summary of the responses..

Platform	Course lecture and email
Title	Computer Desktop and Workspace
Task (pic, survey open closed # Qs)	1. Take two pictures and submit via email 2. Answer the question “Why did you choose to participate?”
Reward/# requested	\$0/200 picture sets (offered to' show pictures and present research goal and conclusions)
Assignment Duration Max work time	n/a
Scheduled run time	7 days
Date and time	1-Sep-10
Assignments Requested	1 from each of the 200 students
Assignments Received	95 picture sets/66 follow-up answers
Assignments Rejected	1 picture set was corrupt, 1 was unusable
Average [work] Time/Hourly Rate	NA/\$0
Completion Time	3 responses arrived after the deadline
Runtime	10 days

Results

From the physical crowd of 200 students 95 responded with the requested picture sets (47.5%). One response had corrupt files, however they were resubmitted upon request. A second submission did not minimize open programs, thus blocking the view of the desktop configuration, and was the only unusable result. Fourteen responses used a screen cap-

ture rather than a picture of the computer desktop; this deviated from the instructions, but did not affect their use in the study. A sample of nine photo pairs is shown in Figure 3.11.

The 95 image pairs provided the means for the lead project designer to visually compare the organization of both digital and physical desktops. The analysis revealed that most pairs did not have obvious similarity in their organizational structure, but they often had a similar aesthetic. Thus, they provided an answer for the hypothesis, though not the desired or expected result.

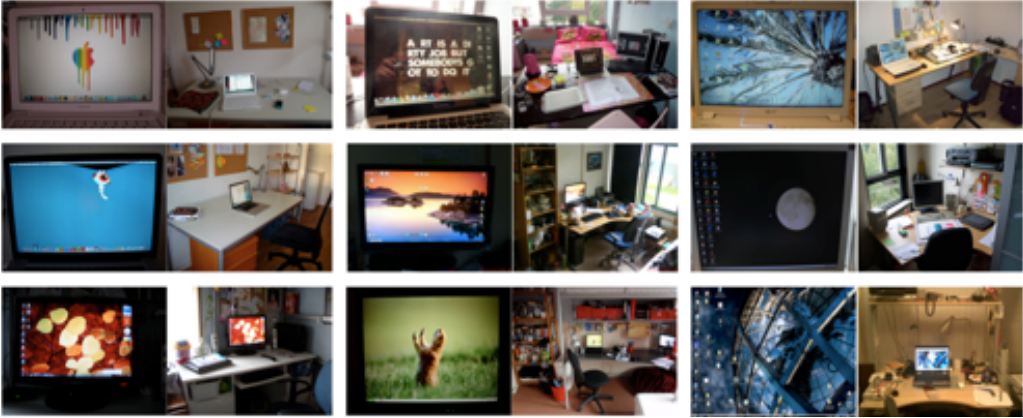


Figure 3.11 A sample of nine of the 95 desktop photo pairs, providing a snapshot of the differences and similarities between the students' computer desktop and their surrounding workspace.

Of the students who submitted pictures, 69.5% (66/95) also responded to the follow-up question asking why they chose to participate. Of these 65% (43/66) gave more than one reason (see figure 3.12).

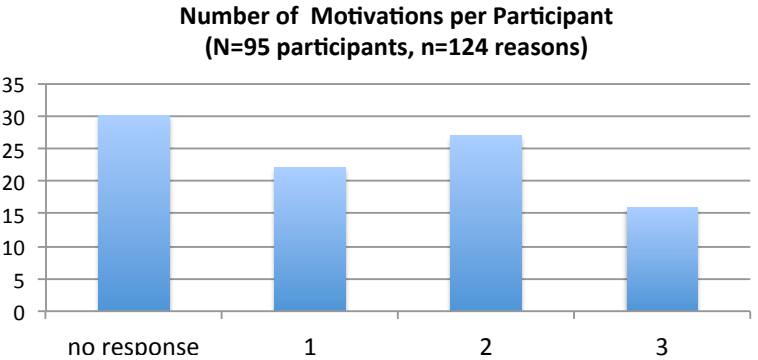


Figure 3.12 The number of reasons reported by each participant for why they chose to participate. Of the people who responded 65% reported more than one motivation.

The students who reported their motivations showed considerable diversity in their reasons for participation (see figure 3.13). Their reasons included: interest or curiosity, supporting research, ease of the task, paying-it-forward, or a false impression that the task was obligatory or a required class assignment.

Figure 3.13 The distribution of motivations reported by participants for why they chose to participate. Note that interest, simplicity, and helping out are the key motivators.

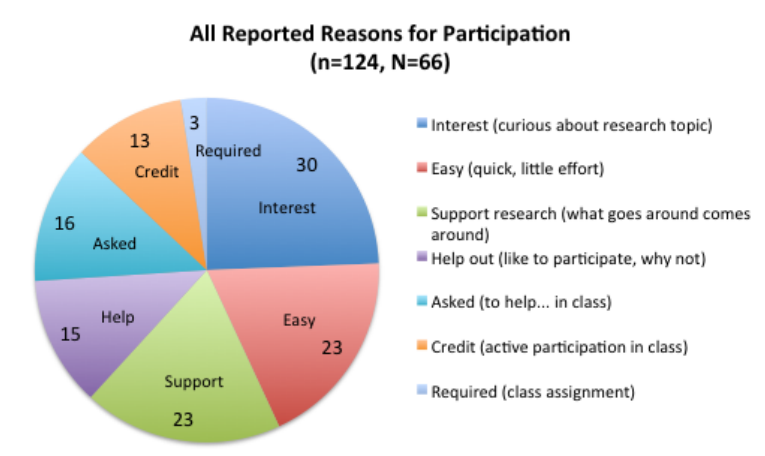
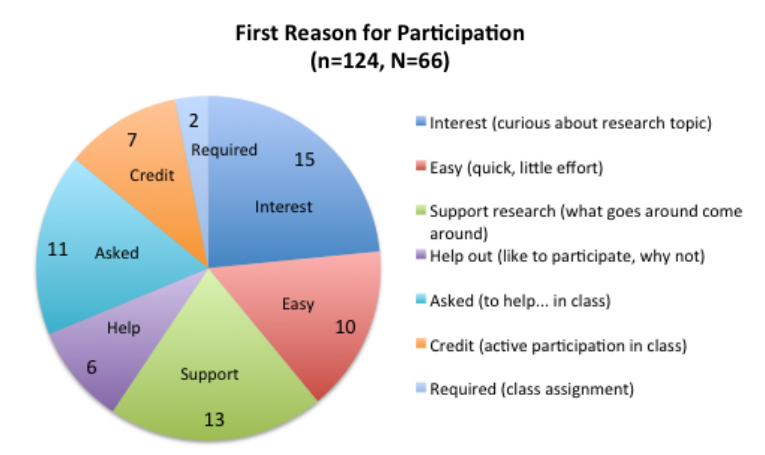


Figure 3.14 The first (possibly primary) motivation reported by each participant reiterates that interest, simplicity, and helping out are the key motivators.



Discussion

The results showed that physical crowds operate under similar principles to the crowds on online crowdsourcing applications, though there are some distinct differences. In this off-line setting, common motivations included: interest, little effort, and helping out. Those motivations align with previous research on reasons for participation in online crowdsourcing. What is missing in this setting is an explicit reward for participation, but this appears to be replaced by a sense of obligation, even though their participation had no relationship to their course grade. This sense of obligation is likely responsible for increased quality (followed instruction and no spam) and a higher response rate.

The complete absence of spam responses and the high quality of the images creates a clear and useful dataset for analysis. Only one participant provided an image that could not be used. This level of quality was

not expected. Past experience with picture submission tasks on MTurk resulted in approximately 10% of the responses being unusable (do not match the request, few images, low resolution). If possible, evoking a sense of obligation may decrease the incidence of spam from online crowdsourcing applications.

Also of interest was the high rate of participation, nearly half (47.5%) of the students responded to the initial task and 69.5% of this first set responded to the unexpected second task (question on participation). In contrast online participation in crowdsourcing often refers to “the 1% of the 1%” (Howe 2008): this is a generalization that 1% of Internet users contribute online and 1% of those participate in crowdsourcing activities. In comparison, a meta-study of email surveys revealed a mean response rate of 34% (SD 15.7%) and that the number of contacts, personalized contacts, and pre-contacts are the primary factors that affect response rates. This study was conducted along similar lines.

The higher response rate of this study appears to be linked to multiple motivational factors since many participants (65%) expressed multiple motivations. The most common reason was “interest in the topic” (24.2%). This coincides with crowdsourcing literature that emphasizes the combination of motivations to activate participation and encourage quality (Bernstein et al. 2010, Kittur et al. 2011, Lahkani & Panetta 2007). The reported motivations suggest that this task found a “sweet spot” for the crowd. Not only was it interesting, it was also perceived to be a “low effort” opportunity to “support research” and to just “help out.” These factors are also reported as motivations in other crowdsourcing activities (Kaufmann et al. 2011, Kittur et al. 2011, Lahkani & Panetta 2007), but they do not account for all of the respondents.

The other predominant motivation reported was a sense of obligation 25.8% (“required” 2.4% + “participation credit” 10.5% + “asked in class” 12.9%). While they were not told it was required nor would they receive course credit, the classroom setting (where completing assignments for a professor is likely) influenced at least a quarter of the participants. This hierarchy or sense of social obligation toward participation does not exist in online platforms.

Overall these results support existing research on motivations in online participation and the activation of multiple motivating factors (easy, fun, interesting, etc.). This study also provides insight on how motiva-

tions may be used to influence respondents, including appealing to the intrinsic motivations of the crowd and leveraging social obligation regardless of where the crowd resides.

Expecting the unexpected, the pictures revealed far more insight into their workspaces and how they lived, emphasizing that these workspaces were not at an office. This bias precluded our ability to compare office workers. A physical crowd logically has some homogenizing/biasing characteristics that are arguably more pronounced than the “online crowd” (they are in the same room for known reasons).

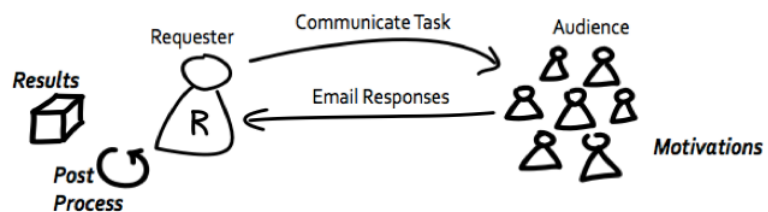
Connection Between Digital and Physical Desktops

The analysis of the submitted picture pairs did not show a clear relationship between the organizational structure of computer desktops and workspaces, though they often had similar aesthetic characteristics. While this was not the desired or expected result, it did provide an answer to the user research question without having to visit the homes of 95 people. We received a unique and surprisingly personal view into the way participants live and work without physically intruding into their homes.

Conclusions for the Framework

The results demonstrate that in general the principles governing motivation and participation in online crowdsourcing activities clearly relate to an off-line crowd. The key difference with engaging a physical crowd is not in the crowd itself, but the removal of a mediating platform that typically acts as a level of separation from the crowd lending the researcher (and the crowd members) anonymity. With a physical crowd, the researcher has a closer connection with participants, increasing the sense of obligation of the crowd to participate and submit work of appropriate quality. The downside to this stronger sense of obligation is a devaluing of individuals’ intrinsic motivations, which are unique motivators on crowdsourcing applications where people elect to participate in tasks that align with their interests and their perceived ability to contribute.

Figure 3.15 Without the support of a mediating platform the requester is left to devise a method of exchange and post processing, and mechanism for explicit/financial reward. The results also show how motivations play a significant role in how and why people respond, showing links between on and offline participation.



The lack of a mediating crowdsourcing application also requires alternate methods to communicate and gather responses from the crowd. In this case the initial announcement was given orally and accompanied by a small printed flyer. The study later relied on email to submit and gather responses from participants. This solution worked well for this crowd whose members were physically available and not associated with a specific crowdsourcing application. On the other hand it was cumbersome for dealing with responses from a crowd of this size. In the future a web-form, photo stream, or shared folder would more efficient.

Study 2: Crowdsourcing Off-Line: Tips for Design Practice

Based on the results and experiences from this study the following tips and lessons learned are shared to help future designers in adopting and making the best use of crowdsourcing in their own design projects.

Tip 1: A physical crowd is motivated similarly to online crowds. The removal of the platform as a mediator increases the connection between crowd and requester altering motivations for participation. This increases the sense of obligation and improves response rate and quality, but offsets the typical intrinsic

motivations that drive self-selection in online crowdsourcing participation.

Tip 2: A physical crowd will inherently have less diversity than an online crowd, which may or may not be desirable.

Tip 3: The lack of a crowdsourcing application creates a need to define alternate means to manage the communications and responses from the crowd. A large number of email responses is cumbersome and inefficient to manage, if not planned for in advance and/or automated in some way.

Framing Crowdsourcing for User Research

The results from the studies in Phase I (Chapter 3) in combination with literature and our initial framework (Chapter 2) provided the insight to develop a conceptual framework of crowdsourcing as a tool for user research (developed below). The framework developed here serves as a foundation for the remainder of the research, providing three key functions. First, it provides the initial connection between current use of and research on crowdsourcing to support a transition into a tool for user research. Making this transition is the first step in answering the main research question and provides a vision of crowdsourcing in a design context. Second, the framework highlights differences in both typical crowdsourcing and common user research processes. These distinctions serve as an impetus for the research questions addressed in the next Chapter. Finally it provides a structure to organize the studies and findings into a cohesive whole.

4.1 Reframing Crowdsourcing for User Research

The framework presented at the end of Chapter 2 provides a clear depiction of the crowdsourcing process. It shows the exchange of task, response, and reward between the crowd and host via the crowdsourcing application as a tool for production. However, it does not reveal how crowdsourcing could function as a tool for user research. Reframing the framework is more than a simple renaming of elements. As experienced by the researcher in phase I, the transition starts with a shift in mindset away from work production and toward an explorative process of accessing users and seeking their insights. These considerations help to reflect on the goals and perspective of user researchers.

The framework is relabeled, shifting our perspective to fit into the language of design and user research. This represents how the actors and their goals differ from production crowdsourcing. As shown in Figure 4.1, four elements in the framework are relabeled: Requester, Crowd, Responses, and Result. In a design context, the role of the *requester* becomes the responsibility of the *designer* or user researcher. The *crowd*

becomes an audience of people and potential *users*, and their *responses* become *user insights*. The end *result* becomes *information* from and about users. These new labels, though generic, allow us to begin thinking about crowdsourcing in this new context as a tool for accessing user information.

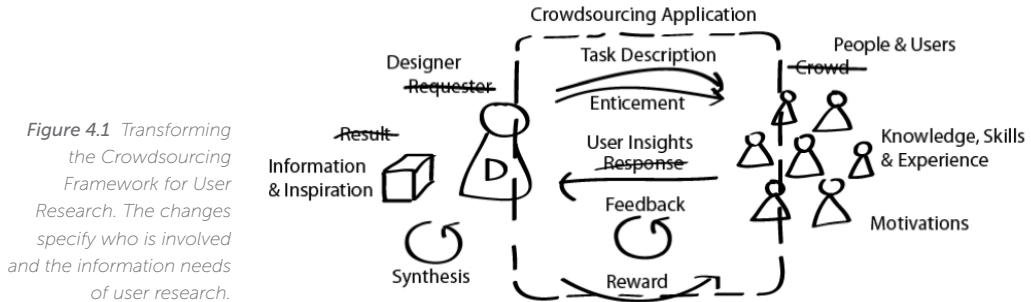
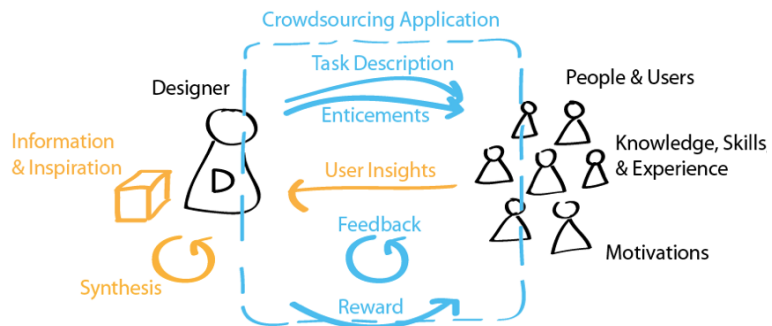


Figure 4.1 Transforming the Crowdsourcing Framework for User Research. The changes specify who is involved and the information needs of user research.

4.2 A Conceptual Framework of Crowdsourcing for User Research

The conceptual framework of crowdsourcing for user research shown in Figure 4.2 shows the result of the preceding transformation. The combination of these changes into a framework provides a view of this new use of crowdsourcing and serves as a foundation for the remainder of our research.

Figure 4.2 The Conceptual Framework, depicting crowdsourcing as a tool for user research in design. The colors highlight the people involved (black), the role of the applications as the interface between designer and users (blue), and the handling of insights from the users (yellow). It is this framework that serves as the foundation for each study and is part of the core of our research.



The transition of crowdsourcing, away from production into a design mindset, highlights several key changes and provides insight into the use of crowdsourcing for user research. First, the *requester* is now identified as *designer*, emphasizing that the designer is in the driver's seat for the process and is creating the task that addresses their unique information needs. Second, the *crowd* is no longer a homogeneous entity; instead it is a group of *people and users*. This distinction recognizes that the crowd is composed of a diverse population of real people who respond to the

tasks. These fundamental changes have a ripple effect throughout the framework, altering what is asked for (*Task Description*), how *users* respond (*User Insights*), and ultimately how their individual responses (*Information and Inspiration*) are viewed and processed (*Synthesis*). While the basic structure of the process appears to remain the same, this transition is not simply a renaming of elements. It is a realization that the transition evokes changes throughout the process that need to be clarified and understood in order to address the goal of this research (Chapter 1.3): apply the benefits of crowdsourcing to user research to bring information from users into the formative stages of design. Figure 4.2 is a visual overview of the changes followed by a description of each element and how they have changed to reflect the application of crowdsourcing as a user research tool for design.

In this transformed framework, each of the elements takes on new characteristics, reflecting the shift in mindset of the overall process toward user research. The list below highlights the changes in each element.

1. Designer (previously the **Requester**): This was renamed to clearly show the placement of the designer (or design researcher) in the central role of the process, and their connection to all the other elements. As the initiator of the entire process, designers select the platform, develop the task, synthesize the insights, and provide feedback and reward. All of this now centers on the designer's information needs. Placing the designer in the role of requester brings him into the information seeking process and emphasizes the desire for a fast and flexible process that fits the designer's process and information needs.

2. Crowdsourcing Application (unchanged): While this is not renamed it is important to note that there are many applications and each is different (e.g. GUI, typical tasks and rewards, size and makeup of crowd). These differences influence the feasibility and appropriateness of a specific platform for different tasks and user populations.

3. People and Users (previously the **Crowd**): The crowd is renamed to emphasize that it is not a single entity, but rather a group of individual people, including members of the desired user group. This emphasizes that we are asking people for their unique insights (no longer seeking task completion) and recognizes that who responds is often a critical component of the insights they provide. The dual label highlights that not all of the people in the crowd will be members of the target user group.

- 4. Knowledge, Skills and Experience** (expanded to include *Experience*): As with any crowdsourcing task, only a subset of the knowledge and skills inherent in the crowd are of interest. For informing the design process, certain experiences may be key to providing relevant information to designers.
- 5. Motivations** (unchanged): Again the people contributing on a crowdsourcing application do not change, but we are likely interested in only a subset that represents a desired user group. Understanding their motivations helps develop the task description and encourages quality inputs, and may also help in reaching the desired user group.
- 6. Task Description** (unchanged): The role of the task description (communicating the task) remains the same. In the new context, the task being requested changes to meet the information needs of the designer.
- 7. Enticements** (unchanged): The role of enticements to encourage appropriate/desirable responses remains the same. The implementation may shift to encourage responses for a specific group of users and to seek the information desired for user research.
- 8. User Insights** (previously *Responses*): As a tool for user research the responses are no longer just task completions or work production. Therefore, they are renamed to distinguish the types of responses that a designer will solicit from users.
- 9. Feedback** (unchanged): The feedback mechanism built into most crowdsourcing applications may serve as a communication link between the designers and people and users. Often it is used to explain why a response did not receive the reward or, in crowdsourcing competitions, to help refine responses.
- 10. Reward** (unchanged): The reward remains unchanged in name and role, but it remains an important consideration that plays a significant role and influences response rate and quality. Balancing reward with intrinsic motivations may also encourage users to provide the rich insights desired by user research.
- 11. Synthesis** (*new element*): This new element was added to show that there is an added level of analysis involved with interpreting inputs from users that goes beyond the typical filtering and consolidation of crowdsourcing responses. Synthesis is the process of translating the user insights into information and inspiration that can be used to influence

the design process. It is this essential process that is both the motivation and the goal of conducting user research in the first place.

12. Information and Inspiration (previously the **Result**): The end result of the process is renamed to reflect the designers' information needs: to inform and inspire the design process based on a better understanding of the users and their needs.

4.3 Role of the Framework

The transition of the crowdsourcing framework provides a foundation for the second phase of research studies (Chapter 5) that seek to explore the value and usefulness of crowdsourcing user research in a design context. In addition, the framework will help organize the findings so they can be useful to designers. The framework already reveals changes in how the elements operate and raises questions about whether crowdsourcing can meet the information needs of designers and fit within a user centered design process. Consequently, the framework as well as the questions it supports becomes the focus and emphasis for the remainder of our research.

4.4 Research Questions

Crowdsourcing literature, as described in Chapter 2, illuminates who participates, why, and the diversity of what they contribute. Existing literature, however, does not address the use of crowdsourcing in the context of design; to reach users and gather insights for the design process. The framework provides a visual map to begin building our knowledge of how crowdsourcing may be applied to user research in design. To focus the second phase of studies and provide designers the desired information, our research questions were derived from designers and design researchers as they expressed concerns about using crowdsourcing. These concerns were captured early in the research process as our research ideas were met with apprehension and questions from designers. These questions and concerns consequently providing a goal that would aid designers in adopting crowdsourcing for user research. Table 4.1 connects the research questions to the conceptual framework and to the designers' questions and skepticism as it relates to crowdsourcing as a user research tool. These research questions serve as the starting point for phase II (studies 3-6) and guide the research exploration of crowdsourcing in a design context. The questions are answered at the end of each respective study and summarized in the beginning of the conclusions (Chapter 7).

Table 4.1 The research questions and their relation to questions and concerns raised by designers during this research, and how they connect to the framework. The phase I studies seek to build our understanding of crowdsourcing in relation to familiar user research activities, the remaining four studies in phase II then explore the value of crowdsourcing for designers.

Study	Research Questions	Designers' Questions/Skepticism	Connection to Framework
1. Crowdsourcing vs. Web Search (Ch 3)	What distinguishes crowdsourcing from web search as a source of user information for design?	"Why wouldn't I just search the internet for user information? How is this any different?"	Distinguish crowdsourcing from web search, and the role of the crowd in producing the result.
2. Crowdsourcing Off-Line (Ch 3)	What distinguishes online crowdsourcing from questioning a known audience (or off-line crowd)?	"Why ask a crowd of online strangers when I have an audience at hand? What differences do I need to consider?"	The role of the application as the connection between designer and crowd, and participant motivations.
3. Online Design Contests to Inform Design (Ch 5)	What can designers gain from hosting a crowdsourcing design competition?	"Why would I outsource my own job to a bunch of amateurs?"	The designers' role and influence throughout the process, and the role of the feedback in improving responses.
4. Formative User Research with MTurk (Ch 5)	What is the value of crowdsourcing as a user research tool in design?	"How should I use crowdsourcing in my design process? Where does it fit?"	The role of crowdsourcing to provide user insights in the design process.
5. Fast and Good (Ch 5)	How fast is crowdsourcing for user research tasks?	"So it's faster than most user research, but how fast? And won't the increased speed ruin quality?"	The role of the crowd's motivations in writing the task description to influence speed and quality to responses.
6. Finding Target Users (Ch 5)	How can crowdsourcing reach people with specific expertise or knowledge?	"How do I get responses from a specific user group, not a crowd?"	Using qualifications to target users and identifying useful responses from the crowd.

In Table 4.1 above, we can distinguish between the studies in phase I (Chapter 3) and phase II (Chapter 5). Phase I strengthened our understanding of what happens and how it is accomplished, while the studies in phase II will explore the value of crowdsourcing for designers. The table above also shows the connection between each study and its link to the conceptual framework. As presented in Chapter 2, the first two studies seek to understand the process and mechanisms of crowdsourcing user research by examining different types of crowdsourcing in relation to activities familiar to designers. The third study looks at what designers can gain from outsourcing design work and gathering information. The fourth study examines the use of crowdsourcing in design projects. The fifth examines the design of the task description to test just how fast user insights can be gathered. Finally the sixth study looks at mechanisms to see if a specific set of users or expertise can be identified and reached from within the crowd.

4.5 Supporting Research Approaches

Although each study uses a different research approach, three methods are used to provide a unified foundation: Action Research, Grounded

Theory, and Design Research. In addition, each study includes a method section that details the specific research approach. To understand the basis of each study it is beneficial to first summarize the three underlying research methodologies.

Action Research

Action research, a qualitative research approach from social science, was selected as the predominant approach. Action research emphasizes the embedded role of the researcher, providing an ideal means to gather data by participating in the decisions and implementation of crowdsourcing as a user research tool. As originally described by Lewin (1946), action research consists of iterative cycles of “planning, action, and fact finding about the result of the action.” Through these cycles, action research seeks to build understanding of a social construct by trying to change it (Brydon-Miller 2003). The act of change and observation of the effects is accomplished through active participation of the researcher within the phenomena or context of interest. This method is sometimes referred to as *participatory action research*, to emphasize the active role of the researchers throughout the process in deriving, selecting, applying, and reflecting on the outcomes of the actions. Integration with practitioners produces outcomes that are more likely to be valid and usable (Argyris 1999).

Through the use of action research, I as the researcher play a role as a member of the design team, assisting in the design process as the host or *requester* of the crowdsourcing activities, and working with the designers (as study participants) to create the crowdsourcing tasks and provide insights into the process, raw outcomes, and application of the results. This separation in roles deviates from pure action research in that the designers are not fully integrated into the research process. However designers assisted in the refinement of the research questions and provided an additional lens to examine the use of crowdsourcing to gather insights from users. The iterative nature of *action research* is applied in the individual crowdsourcing activities conducted within and across each study. Each iteration contributes insights that are incorporated into later iterations as we learn more about the application of crowdsourcing as a tool for user research. The added breadth aims to provide results that support user research in design projects with a wide variety of topics and information needs.

Grounded Theory

Grounded theory is a research method that seeks to develop a new theory that identifies and then describes the general features of a phenomenon. The explorative nature of grounded theory provides a means to experience crowdsourcing with an eye to uncovering the unexpected. This is contrary to scientific method, which uses a carefully designed experiment to test a predefined theory or hypothesis. Instead, a theory is derived through inductive analysis of empirical (observation and experience) data from multiple sources (Glaser & Strauss 1967). Theory formation uses coding and categorization analysis techniques that do not rely on existing theory or predefined categories, allowing the features of a new theory to emerge from the data. The grounded theory approach asks the researcher to be open to new discoveries (Corbin & Strauss 1990), while limiting theoretical preconceptions. Uncovering emergent themes enable us to develop a framework of crowdsourcing as a user research tool. The discoveries are incorporated as individual data points to inform the framework.

Design Research

This research also borrows from the different approaches to design research, as distinguished by Horvath (2007): Research in Design Context, Design Inclusive Research, and Practice-based Design Research. *Research in Design Context*, refers to the application of scientific research methods to gain insight into various aspects of the design process. What distinguishes this approach from the classical empirical approaches is that the research is not de-contextualized; the inquiries are conducted specifically to gain insights into the phenomena of a specific design context, with all the clutter and distractions of reality intact. We recognize that the context of each study (topic, designers, the crowdsourcing application, and numerous uncontrolled factors) cannot be separated from the research and influences the outcomes and their interpretation.

Design Inclusive Research combines design practice and fundamental research. At this intersection, the design process is considered to be a “knowledge synthesis process” conducted within the research setting. In this way, the process of design is used to provide knowledge to test the research hypothesis. This resembles the

Research Through Design approach where the act of designing and creating solutions is in itself a knowledge generation process (Archer 1995, Stappers 2007, Sleeswijk Visser 2009). The iterative process of design reevaluates assumptions and refines the design as knowledge is gained.

I did not study the act of design itself; instead we captured the process and decision-making in creating and completing the crowdsourcing activities. It is the refinement of this process through experience that reveals knowledge about crowdsourcing in this new paradigm. The iterations then refine the framework and inform later activities.

Practice-based Design Research positions the designer as an observer or researcher. This approach, appropriated from research in fine arts, thoroughly plans the design process to create a work of art with a certain objective. While this approach does not follow classical research methodologies, the designer thoroughly documents the design iterations, thus generating knowledge into how (or if) the design objectives were met. This cycle of documenting the actions and decisions made while trying to reach an objective is reminiscent of an action research approach that plays a prominent role throughout our research.

By combining select aspects of these broader research methods, I was able to develop an exploratory approach to examine the process and mechanisms of crowdsourcing. This approach provided a means to capture the insights necessary to address the research questions while conducting the research in realistic settings and in conjunction with design projects.

4.6 Role of the Researcher

In using an action research paradigm, as the researcher, I embedded myself in the crowdsourcing process and as a member of the design team (where possible). This decision provided greater opportunity to observe and collect data throughout the entire process. The user information collected during the studies was often used in design projects. These two influences induced a desire for success, a bias common in participatory research. Persistent awareness of this influence helps mitigate its effects during collection and the analysis of each study. The act of balancing involvement with observation of self, team, and outcomes was a unique and informative challenge. The specific research methods and data sources for each study are described in detail just prior to their results (Chapters 3 and 5). Insights gained through my perspective as a researcher and playing different roles in the process added to the understanding of crowdsourcing user research.

Transitioning the crowdsourcing framework for user research provided a foundation to organize the research questions and findings. The

adoption of an exploratory research approach that combined aspects of action research, grounded theory, and design research provided the opportunity to participate in the research to gather insights and seek out the unexpected. In the next chapter, four studies seek to answer the main research question: How can designers combine the benefits of crowdsourcing with user research to bring information from users into the formative stages of the design process? The studies examine the role of the designer, the ability to provide user insights, task design in relation to speed and quality of responses, and how crowdsourcing can reach target users.

CHAPTER 5 Phase II - Studies

In order to answer the research questions highlighted in the preceding chapter, four studies were conducted. This chapter details the four main studies designed to understand how designers can use crowdsourcing as a formative user research tool. The studies explore and clarify different aspects of the crowdsourcing framework constructed in Chapter 4. Not only do the studies inform the framework, they also address common perceptions, misconceptions, and questions expressed by designers when considering the use of crowdsourcing as tool for conducting user research (Table 5.1). Each study focuses on different aspects of the processes, interactions, outcomes, and fit in the design process. In addition to addressing the key questions and concerns, the explorative research approach (as described at the end of Chapter 4) allows examination of the research questions from the perspective of the participating designers.

The studies are presented in chronological order; as the experiences gained through each helped inform the next. At the end of each study the results were fed back into the framework in Chapter 4. We conclude the presentation of each study with a list of tips or lessons learned to guide designers in their future use of crowdsourcing.

Table 5.1 (next page) provides an overview of all six studies (including the first two formative studies presented in Chapter 3). The table expands on Table 4.1 shows the main research question, designers' questions related to their skepticism, connections to the framework, adding the context of each study and the related publications. Chapter 6 then provides a general discussion of the implication of the results for both research and design, and closes with a discussion of limitations.

Table 5.1 An overview of the two research phases and the six studies (expansion of Figure 4.1)

Study	Research Questions	Designers' Questions/Skepticism	Connection to Framework	Context	Related Publications
1. Crowdsourcing vs. Web Search (Ch 3)	What distinguishes crowdsourcing from web search as a source of user information for design?	"Why wouldn't I just search the internet for user information? How is this any different?"	Distinguish crowdsourcing from web search, and the role of the crowd in producing the result.	4 sets of 40 images were gathered using, MTurk, Google Images, Flickr (x2) iStockPhoto.	Tidball, Mulder, & Stappers (2011) Inspiring Design: Exploring Online Sources of User generated Information. <i>Proc. CHI Sparks 2011</i> , Arnhem, NL.
2. Crowdsourcing Off-Line (Ch 3)	What distinguishes online crowdsourcing from questioning a known audience (or off-line crowd)?	"Why ask a crowd of online strangers when I have an audience at hand? What differences do I need to consider?"	The role of the application as the connection between designer and crowd, and participant motivations.	1 off-line crowdsourcing task using a course of 200 students as the crowd.	
3. Online Design Contests to Inform Design (Ch 5)	What can designers gain from hosting a crowdsourcing design competition?	"Why would I outsource my own job to a bunch of amateurs?"	The designers' role and influence throughout the process, and the role of the feedback in improving responses.	1 logo design contest, conducted by a student design organization.	Tidball, Mulder, & Stappers (2011) Online Design Contests: A Network of Inspiration for Designers. <i>Proc. IASDR, 4th Intl. Conf.</i> , Delft, NL.
4. Formative User Research with MTurk (Ch 5)	What is the value of crowdsourcing as a user research tool in design?	"How should I use crowdsourcing in my design process? Where does it fit?"	The role of crowdsourcing to provide user insights in the design process.	7 design projects hosted 27 crowdsourcing tasks on MTurk	Tidball, Mulder, & Stappers (2017) Crowdsourcing as an On-demand User-research Tool. Submitted to TOCHI
5. Fast and Good (Ch 5)	How fast is crowdsourcing for user research tasks?	"So it's faster than most user research, but how fast? And won't the increased speed ruin quality?"	The role of the crowd's motivations in writing the task description to influence speed and quality to responses.	6 crowdsourcing tasks hosted on MTurk during design workshops.	Tidball, Mulder, & Stappers (2017) Crowdsourcing as an On-demand User-research Tool. Submitted to TOCHI
6. Finding Target Users (Ch 5)	How can crowdsourcing reach people with specific expertise or knowledge?	"How do I get responses from a specific user group, not a crowd?"	Using qualifications to target users and identifying useful responses from the crowd.	6 crowdsourcing tasks on MTurk, conducted in connection with two established research projects.	Tidball, Mulder, & Stappers (2017) Crowdsourcing as an On-demand User-research Tool. Submitted to TOCHI

5.1 - Study #3: Online Design Contests to Inform Design

Designers' Skepticism: "Why would I outsource my own job to a bunch of amateurs?"

The preceding two formative studies (Chapter 3) highlighted how crowdsourcing, in the form of micro-tasks (both online and off-line), can be used to gather insights from users to inform design. In a similar vein, the current study looks at how design contests intended to generate solutions can be used as a source of information for designers. By examining the effects of this role reversal – from designer (creating solutions) to client (communicating the goal to other designers) – I expect to illuminate how crowdsourcing can inform the design process.

To explore these ideas, this study observed two masters level design students and a staff graphic designer as they collaboratively hosted a logo design contest on hatchwise.com. Rather than selecting a winner, the aim of the researcher during the contest was to observe how the designers deal with other peoples' solutions as a source of inspiration and an opportunity to broaden their view of potential solutions (see Figure 5.1). Due to the contentious nature of this topic, specific attention is given to potential ethical concerns and this is discussed at the end of this study.

Research Questions

- What can designers gain from hosting a crowdsourcing design competition?
- *What is the role of feedback (beyond "thanks") in communicating with contributors?*
- *What value do the submitted designs contribute to the design process?*
- *What are the ethical considerations of using a contest to inform design?*

The growing popularity of online design contests (a common form of crowdsourcing) often evokes negative reactions from veteran designers. Concerns about being replaced with cheap amateur design work, reduced job opportunities, and a watering down or devaluing of their skills and profession are common (Grefe 2011, ycombinator.com 2013). However, research is inconclusive on the impact online contests have on the design profession. Research literature shows that contributors generally have some training or qualification and that they produce quality de-

This study is based on the following publications:

Tidball, B., Mulder, I. and Stappers, P.J. (2011). Online Design Contests: A Network of Inspiration for Designers. In *Proceedings of IASDR, 4th International Conference*, Delft, NL.

Tidball, B., Mulder, I. and Stappers, P.J. (2017). Crowdsourcing As An On-demand User Research Tool. Submitted to *ACM Transactions on Computer-Human Interactions*.

signs. Rather than pick sides, this study examines how a designer could put these contests to use to inform their own design projects.

As shown in Figure 5.1, specific attention is given to the communication elements of the process and interactions between participants (the design brief and feedback loop) and potential sources of information (the submissions and synthesis of the results).

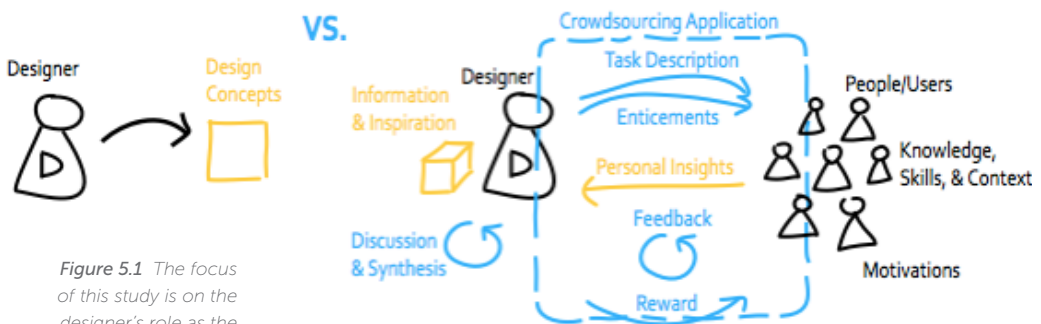


Figure 5.1 The focus of this study is on the designer's role as the host – rather than the contributor in an online design contest – as an opportunity to gain insight and broaden their view of potential solutions. It also explores their communication with contributing designers through feedback as an additional means of interaction with the contributors.

Context

This study was conducted in connection with the development of a new interaction design community within the Industrial Design Engineering Faculty at Delft University of Technology. This community planned to host “4 Interaction,” an online magazine dedicated to interaction design. This magazine and community needed a logo to identify events, publications, and projects associated with it. To facilitate the process of creating a logo, the faculty offered the student organization an opportunity to host an online design competition to quickly generate a wide variety of proposals. The contest was then studied to develop an understanding of the communication and interactions with contributors and the potential benefits for informing the design process. Hatchwise.com was selected because it is a popular, active, and inexpensive crowdsourcing application that specializes in logo design contests.

Method

To explore the research questions, this study focuses on observing two aspects of a crowdsourcing design contest as a source of user information for design:

- The benefits of conducting a design contest as a source of information and inspiration for designers, including what was learned from hosting a contest and how designers use the results.
- The interactions and communication between contest hosts and the contributors, as well as the impact they had on the contest and its output.

Data Collection

The data for this case study were collected from three sources: the contest website, researcher observations, and student experience papers. The contest website provided an automated and structured way to capture information. This information included the design brief, the number of contributors, their submissions (logo designs), feedback from the contest hosts, and comments between contributors. In addition, as lead researcher, I acted as advisor to the two students who hosted the online logo contest. In this role, the researcher provided “how to” support and was able to observe and document their activities and discussions. The staff graphic designer also assisted with common terminology and recommendations for clarity. The researcher observations primarily captured preparation activities and discussion of the results. To document the contest from the hosts’ perspectives (and to promote a learning experience) each student host wrote a paper providing personal insights and reflections on their experiences. In combination, these sources of information provided insights helpful in developing answers to our research questions.

Setup of Crowdsourcing Activity

The contest was setup on hatchwise.com following the built-in logo design options. Figure 5.2 shows the setup parameters as displayed by the initial posting and Figure 5.3 shows the full design brief (next page)

Current Projects	Started On	Time Left	Prize	Prize Prepaid?	Entries Submitted	New Entries?	New Comments?	Comments	Action
Logo Design Contest for "4 Interaction" Design a Logo for a Design Community	2010-12-03	9 days 23 hours 59 minutes	\$100.00	Yes	0			0	View Contest Upgrade Contest

Figure 5.2 The posting of the design contest on the hatchwise.com website, as it appeared on the website to potential designers.

The contest description was designed to encourage variety while staying within desired boundaries “You should take a look at the TU Delft website www.tudelft.nl. The logo you create should feel like it is part of the same “family.” For reference the blue color is Process Cyan (RGB 0/166/214).”

Results

For \$129, the contest developed 120 logo designs (see Figure 5.4) submitted by 32 contributors (approximately a \$1 each). While the contest ran for 10 days, administering the contest involved a little more than six hours, divided evenly among preparation, observation and feedback, and synthesis. For this limited effort and expense the student hosts and researchers were impressed with the variety, and quality of the submissions.

Design Brief:

What is the full name as you want it on your logo?

"4 Interaction"

Do you want any slogans or taglines on your logo?

No

What service or product do you provide?

4 Interaction is an online magazine, focused on Design for Interaction (Dfi). It is hosted by students and faculty of Industrial Design Engineering. The logo will identify events, publications, and projects of interest to the Dfi community.

What is your target audience?

Our audience is anyone involved/interested in Design for Interaction: professors, students, researchers and designers. The online magazine is about people and products, and we hope to be read by professionals and practitioners outside the university.

What style would you like your logo to be? Would you like unique text, an icon with the name included in it, an icon with the name next to it, an illustration, or a different style?

We would love to have it as combination between an icon and a unique text, but feel free to experiment with it and show your creativity.

Do you want any specific ideas implemented into your logos?

You should take a look at the TU Delft website www.tudelft.nl. The logo you create should feel like it is part of the same "family". For reference the blue color is Process Cyan (RGB 0/166/214).

What should come to a customers mind when they see your logo?

Simplicity and cleverness. We want our audience to relate this community of Design for Interaction.

Is there anything you would not like to see in your logo?

We don't want a big mix of colors and we are not looking for extra decoration.

How will the logo be used?

The logo will primarily be used for an online magazine, but will also be used in print advertisements, publications and other activities organized by the community.

Do you have any files or images pertaining to your contest that you want the designers to view?

www.tudelft.nl

What format do you want your logo to be designed in (.ai, psd, corel etc..)

Vector, Adobe Illustrator (.ai)

Figure 5.3 The Design Brief: the details of the task description as seen by contributing designers.



Figure 5.4 The 97 logos contributed to the contest (not including the 23 of 120 submissions that were withdrawn during the contest).

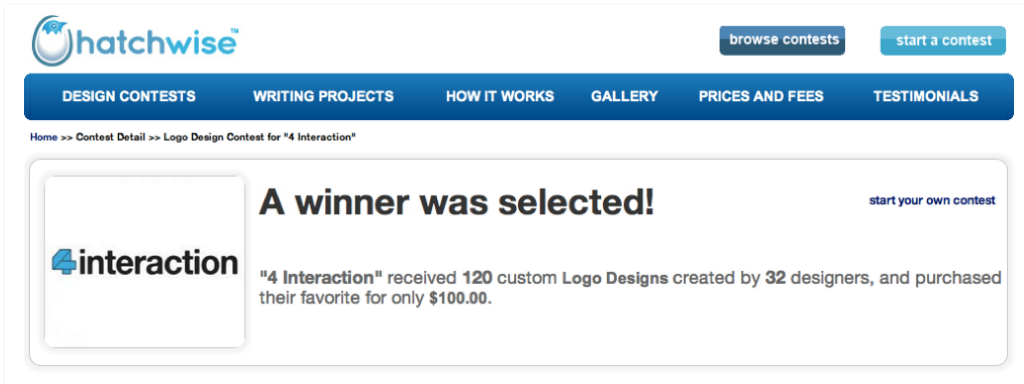
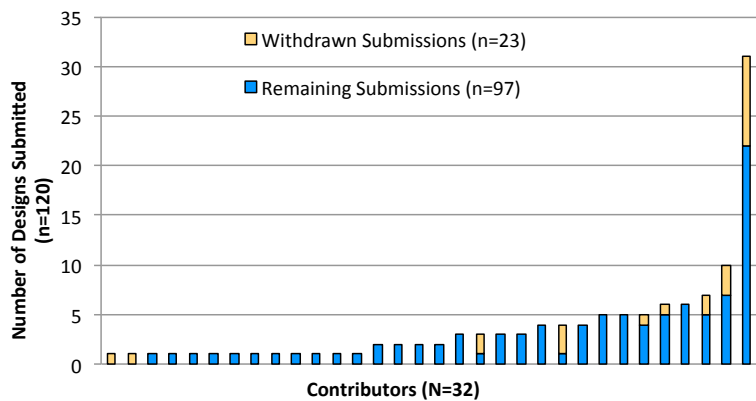


Figure 5.5 The summary conclusion of the contest once a winner had been selected.

Figure 5.6 shows the number of designs submitted by each contributor, including designs withdrawn prior to the close of the contest. Contributors occasionally withdrew submissions after submitting a revision (20 withdrawn designs), two additional designs were withdrawn without submitting a redesign, and one was withdrawn due to accusations of “copying” from another contributor, captured in the general contest comments. The designers were surprised by the commitment of some of the contributors, demonstrated by the number of designs they submitted, especially one contributor who submitted 31 logos (see far right of Figure 5.6).

Figure 5.6 Number of designs submitted ($n=120$), and withdrawn ($n=23$) per contributor ($N=32$). Many designers submitted only one design, and only 5 designers submitted more than 5 designs. Withdrawn designs were usually due to submission of an improved version.



After receiving feedback from the contest host, 11 contributors provided 21 comments of appreciation for the feedback received for their efforts

*"Thanks for the feedback, here are changes, what do you think?
There will be more versions if you want..."*

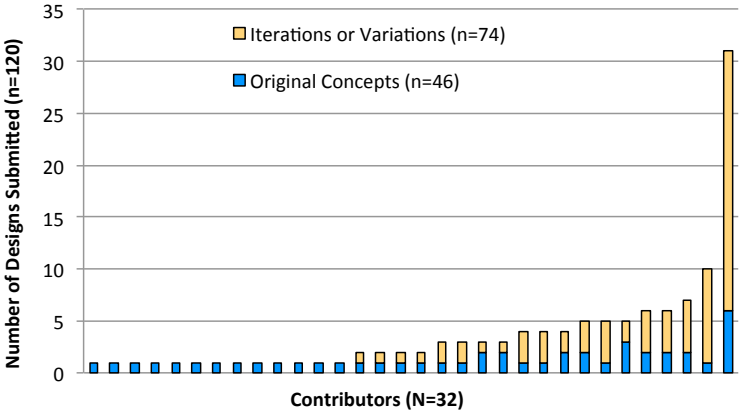
In addition four comments from contributors provided details or described their intention/vision of a particular logo.

"I purposely chose the letter N... what I did is a roman numeral of 4, IV [to make the N]".

While it is not possible to precisely quantify, a little more than half of the iterative logo submissions appear to be responses to specific feedback, while the others are variations on an idea submitted as alternatives.

Figure 5.7 shows the distribution of original design concepts (46 logo designs) and the additional iterative logo designs (74) per contributor. The high number of iterative designs highlights a favorable response to feedback, since contributors used additional effort in refining and resubmitting their designs. Figure 5.7 also shows that only two contributors produced three or more original concepts, indicating a relationship between the number of original concepts and the number of participating contributors.

Figure 5.7 The distribution of original designs (blue) in relation to iterations (orange) submitted per contributor (including withdrawn submissions). The diversity in the designs comes from the number of designers participating.



Observations

In addition to the results from the contest itself, researcher observations and the student reflection papers revealed three noteworthy developments during the design contest. The development of the design brief, the feedback process, and the discussion of the results all provided insight into the benefits of a design contest.

Initial observations noted challenges as the student hosts developed the task description (design brief). Their initial design brief, while clear, provided considerable detail and was highly constraining, reflecting specific ideas from their own initial design attempts. After this was pointed out and the guidance on the contest website was reviewed, they simplified the brief to clarify their need at a more strategic level. Avoiding stifling the contributor's creativity was appropriate since the team was seeking inspiration and diverse designs. This indirect step in the process encouraged the contest hosts to more clearly identify and discuss their information needs. This in turn led them to reflect on and refine their project goals in a manner that could be shared with others. In the end, the brief described the intended use and audience and included a statement that it should be in the "family" of the university's logo (as shown previously in Figure 5.3).

The second notable observation occurred as the hosts began providing feedback to contributors during the contest. They initially expressed apprehension about providing feedback, but with a little encouragement and the guidance of the website, they provided thanks and constructive criticism for most of the submissions. The students later confirmed their apprehension that as design students, "we are used to receiving not giving criticism." They also said that once they started commenting and received appreciation for their feedback, it became easier. These observations were reiterated in their reflection papers. The student hosts later stated that providing feedback strengthened their confidence in their own design knowledge and skills, and helped them think critically about the vision they had for the end result of the project.

The third observation of interest occurred at the end of the contest, when the design team (student hosts, researcher, and a staff graphic designer) met to discuss the results and select a winner. The meeting began with all of the submissions on the table and the student hosts and the staff designer identifying their five favorite designs and describing what they liked about each selected design. The limited overlap in their preferences made it difficult for the team to reach a consensus on selecting a winner, but the discussion enabled and evoked by the submissions proved to be highly beneficial. The large and diverse collection of options provided the team the content and freedom to candidly discuss the branding of their interaction design community and what role the logo can and should play. In the end they did not use the winning design, but the process and resulting discussion provided a wider view of the possibilities, implications, and considerations for developing their logo.

Discussion

The results above address the research questions for this study and provide insights into the interaction between host and contributors, as well as the benefits of the process itself. Everyone involved was impressed by the quantity, quality, and diversity of the designs, especially considering the limited effort and expense of hosting the contest (10 days to run, 6 hours of work, and \$129). The hosts received 46 original designs for $\pm\$2.80$ each.

Communication and interactions with the contributing designers occurred in two primary ways: the design brief and feedback during the contest. As described, the design brief encouraged careful consideration to balance a clear statement of the desired outcome without unnecessarily constraining the skills and creativity of the contributors. The wording of the brief appeared to be key in enticing contributors to participate for motivations beyond a financial reward (interest in a particular challenge or topic). These factors are related to the quantity of original designs submitted: more contributors, more designs.

In addition to providing a communication link and motivation for participation, the open discussion and considerations uncovered during the refinement of the design brief provided additional benefits for the hosts and the design team. While it was difficult to give up control and allow the contributors to present their own ideas, creating the design brief helped clarify what they wanted and needed from their logo beyond their initial ideas. Refining the brief in order to balance needs without stifling creativity forced the hosts to think critically about their desired outcome. This process of simplifying the instructions, eliminating extraneous details, and communicating their desires encouraged fresh insights and a commitment to what considerations were of central importance to their project.

Feedback provided a second communication channel between participants. It was used to thank contributors for their efforts, provide constructive comments, and encourage refinement of their concepts. Their comments were rewarded with re-designs and thanks from 21 of the 32 contributors. These activities confirm the ability of the contest host to effectively motivate the outcome (designs and redesigns). As demonstrated by the positive responses from contributors, this feedback was appreciated and corroborates its value as added motivation for participation, in the form of praise, recognition, and skill development. A few

comments also included details and vision behind some designs. While not utilized in this study, there is a future opportunity to seek and use the additional insights from the contributing designers.

The process of providing feedback as the designs were submitted also proved beneficial. By taking the time to give constructive criticism, the hosts were forced to individually consider how each submission related to their needs and preferences. Unexpectedly it also developed the confidence of the students who were accustomed to receiving, not giving, criticism. By providing feedback they became aware and more confident of their own knowledge and design skills.

In the end it was not the logos themselves that were most beneficial, but the discussion they evoked among the design team. The number, diversity, and anonymity of the custom set of design possibilities provided the content and freedom to address many issues concerning the design: what the design should say about the magazine and community, what constituted a good logo for gaining recognition, and the impact the logo would have on the design of the web-magazine and vice versa. This discussion gave the students a much broader understanding of the solution space and a more informed view of how to reach their objective.

Ethical Considerations

The approach described in this paper has some complications that need to be considered as research and use of online design contests continues. The ongoing discussion about the rights and responsibilities of both contributors and hosts includes concerns of fair compensation, inappropriate and possibly illegal tasks, intellectual property rights, and information privacy (von Ahn et al. 2008; Lakhani & Panetta, 2007; Schmidt, 2010). If the non-winners contribute substantially to the outcome of the process, do they deserve more recognition or remuneration than the “constructive feedback” they receive now? Or is the continued activity of these online contests evidence that contributors receive sufficient fulfillment and reward? This problem is not new. In traditional contests, the spinoff value of the multitude of ideas generated has been an important – if downplayed – motivator for organizations to hold contests. With the increasing popularity and visibility of online crowd-sourced competitions, ethical concerns deserve renewed attention and study.

Conclusions for the Framework

The results of the current study contribute three key insights for the framework (Chapter 4) and support the value of a design competition as a source of information for the design process. First, the development of the design brief (task description), plays a key role as the primary communication to the crowd. Building the brief provides the need and structure to consciously consider and clarify project goals. This includes considerations to attract and motivate participation. To do this well encourages the design team to clearly define their goal along with their information needs and desires, which can help ensure the whole team is on the same page, from the beginning of the project. Second, the feedback mechanism built into crowdsourcing applications is customarily limited to reward or praise. In this study, actively providing feedback on the submitted designs resulted in numerous refinements, but had limited value in seeking new designs or information for the design process. The feedback loop primarily provided opportunities to reward, develop, and refine the skills of both the host and contributors. Finally, the synthesis of the large and diverse set of custom design solutions evoked rich discussions among the designers. This was in contrast to observational and participatory techniques where the individual(s) involved in the user research derive their own inferences before data collection is complete; influencing what information is presented and processed to the rest of the design team. In this case the anonymity of the designs allows for candid discussion that would not likely occur if the contributor were present. This study informed a beneficial discussion focused on the strategic implications of design elements and the influence on future website and print publications.

Figure 5.8 The framework highlighting four elements that were central to the results of this study. The development of task description helped the designers critically consider what their design goals were. Recognition that the crowd was comprised of graphic designers, the use of the feedback loop to encourage refinement (but not new designs), and the freedom they had during the discussion and synthesis of the results were of key benefit to their design process.



The benefits of this study also included insights and considerations into other crowdsourcing activities. It highlighted the benefits of communication among all participants, and the value of different types of mo-

tivation. Equally notable are the often-ignored byproducts of hosting a crowdsourcing activity: in developing the task description, engaging in feedback, and synthesis of the results. This suggests the value of design contests and crowdsourcing activities may go beyond accessing external talent by providing a framework for processing and discussing information.

Study 3: Online Design Contests - Tips for Designers

Based on the results and experiences from this study the following tips and lessons learned are shared to help future designers in adopting and making the best use of crowdsourcing in their own design projects.

Tip 1: The task description (design brief) can be used to promote or restrict the diversity of the submitted designs while still focusing on your design goal. More importantly, carefully crafting the brief affords an opportunity to clarify and refine your information needs and design goals.

Tip 2: The feedback loop is beneficial for refinement and as appreciation and reward mechanism. It is likely possible but would take concert-

ed effort to use it as a communication channel to gather additional information from the designers.

Tip 3: The contest can be a great source of content for critical and candid discussion within the design team without having to worry about insulting or critical comments about the design efforts of a team member. It also facilitates strategic level considerations and discussion among the design team.

Tip 4: Selecting an application that allows for flexible reward structures (\$ for each submission + \$\$ for the top three) helps to address some of the ethical concerns of rewarding all contributors.

5.2 - Study #4: Formative User research with MTurk

Designers' Skepticism: "How should I use crowdsourcing in my design process? Where does it fit?"

The fourth study examines the value and fit of crowdsourcing as a tool for gathering formative user inputs for the design process. The preceding studies looked at specific types of crowdsourcing and their potential for gathering information from users. In contrast, this study embedded crowdsourcing as a user research tool in seven student design projects to observe how it works. Designer students were given the freedom to employ crowdsourcing as they saw fit for their projects and information needs. The research questions below help guide the current study toward its goals to determine if MTurk is a valuable tool for formative user research, to see how it integrates into the design process, and to provide recommendations for future use.

Research Questions

- What is the value of crowdsourcing as a user research tool in design?
- *What are the strengths and weaknesses of crowdsourcing within a design project?*
- *How does crowdsourcing integrate with or detract from other user research?*
- *What are the recommendations for future use?*

Figure 5.9 (next page) provides a simple visualization probing the question of where – if at all – crowdsourcing belongs in the chaos of the design process. More critical than its fit in the process is the question of the value it can provide as a means to bring user insights into design decisions.

This study is based on the following publications:

Tidball, B., Mulder, I. and Stappers, P.J. (2017).

Crowdsourcing As An On-demand User Research Tool. Submitted to *ACM Transactions on Computer-Human Interactions*.

Tidball, B., Mulder, I. and Stappers P.J. (2011).

Online Design Contests: A Network of Inspiration for Designers. In *Proceedings of IASDR, 4th International Conference, Delft, NL*.

Figure 5.9 Does crowdsourcing provide user insights for the formative stages of the design process? How does crowdsourcing fit in the chaos of the design process (Design Squiggle, Newman 2006)?



Motivation

An examination of crowdsourcing in design projects is a critical element in determining how and where it can be used to inform the design process. In contrast to the preceding studies that explore the inner workings of the crowdsourcing framework, this study examines how the framework as a whole fits into the design process. For those well versed in crowdsourcing as a means of production, the results largely reinforce previous findings from other contexts, and add insights necessary for crowdsourcing to become part of the user research arsenal. To take that next step, this study aims to examine the feasibility, strengths, and limitations of such an approach.

Context

This study was conducted in conjunction with seven student design projects. Student projects were selected because they are relatively accessible and observable, they provide diverse project topics, and they have constrained time lines. The students, as relative newcomers to design and user research, can also provide a less biased (though somewhat naïve) assessment of crowdsourcing as a novel approach in contrast to other methods. There are, however, limitations to this choice. The approach lacks the depth of design knowledge and industry experience that comes from professionals. Students lack the depth of design knowledge and industry experience that comes with time in a career. Additionally, our findings in the controlled class project environment do not automatically translate to industry. The expectation is that any advantages or shortcomings identified by the students, despite their limited experience, can serve as a foundation for experienced designers and user researchers to build upon without overlooking the needs of novice researchers.

Participating Courses

The three courses involved in the project are the Joint Master Project (4 projects), Exploring Interactions (2 projects), and a Master's Graduation Project (1 project). These particular courses were selected because the projects emphasize user research, especially for the purpose of informing the early or formative stages of the design process. In addition, the courses required each project to conduct multiple forms of user research, ensuring crowdsourcing would not be the only method used, and allowing for later comparison. Finally, all the topics had a sponsor agency or corporate client, providing some sense of "real world" project requirements. Table 5.2 shows the seven student design projects who volunteered to participate and followed through to the completion of this research project. Five other projects were initially involved, but dropped out of the research primarily due to time constraints and other end of term course obligations, and one project simply stopped communicating.

Table 5.2 An overview of the seven project topics, sponsors, and associated courses included in this study.

Project Topics (Project Sponsor)	Course Name (team size)
Furniture Concepts for Future Work Environments (High-end Contemporary Furniture Manufacturer)	Joint Masters Project (1 semester, 15 EC course, 6 person team)
Airport Retail - Premium Products for Women (International Duty Free Retailer)	Joint Masters Project (1 semester, 15 EC course, 6 person team)
In-Flight Experience for Chinese Passengers (Major World Airline)	Joint Masters Project (1 semester, 15 EC course, 6 person team)
Market Entry Plan for Car-Sharing Services (European Car Manufacturer)	Joint Masters Project (1 semester, 15 EC course, 6 person team)
Insurance Policy Design , for Gen-Y Customers (International Insurance Conglomerate)	Exploring Interactions (1 semester, 9 EC course, individual project)
eHealth for Digital Natives (European Healthcare Provider)	Exploring Interactions (1 semester, 9 EC course, individual project)
Stimulate Contact between Neighbors (Regional Municipality & Urban Design Agency)	Masters Graduation Project (1 Semester, 30 EC, Independent Design Project)

The Joint Master Project is one of the final courses in the master's program. In this 15 credit course, projects of six master's students are assigned a client and follow a design process from inception through several iterations, resulting in a design proposal for the client. The second course, Exploring Interactions is a 9 credit first year master specific design course. In Exploring Interactions, students are introduced to user research and complete a self-selected collaborative design project. The third course is the 230 credit Master's Graduation Project, a one semester solo design project that is the culmination of their master's program.

This project is a collaborative effort with the sponsoring organization, and ends with a final design that is supported throughout the process by user research and has gone through several iterations.

Method

Amazon's Mechanical Turk (MTurk) was selected as the crowdsourcing application for this research because it has a large, active population and a flexible interface. The micro-task structure of MTurk may limit the length or complexity of what can be requested, but this was an acceptable limitation in contrast to other applications whose specializations allow for greater complexity but limit crowd diversity and task flexibility.

Role of the Researcher

In keeping with the action research method, as the researcher, I imbedded myself onto the student design projects as a mentor (in addition to the mentoring of the professor and corporate sponsor). In this role I was able to participate in the projects and observe the students' user research activities. As project and crowdsourcing mentor, I provided an initial overview of how crowdsourcing works, using the framework presented in the previous chapters. The students then developed their own tasks and questionnaires. They received limited guidance in task design, I only pointed out typos or confusing instructions and questions (allowing them to rectify if they chose), and I helped them post their tasks on MTurk. The latter occasionally required limited HTML programming to customize formatting not provided by the existing templates on MTurk, or to customize the labels in the results table. (Note: It is recognized that the API on MTurk allows extensive flexibility and customization, but this study intentionally limited the use of coding to make the approach available to a wider audience.) Finally, I made suggestions for task reward and timing parameters, to align with common practice for posting survey type tasks on MTurk.

Process

Participation of the student design projects was solicited from the three courses described above during a five-minute presentation at the beginning of the semester's first lesson. The projects that expressed initial interest in using crowdsourcing were asked to participate in the study after explaining the research commitment for this research study.

Once the students were familiar with their project topics, they each scheduled one or two meetings with the crowdsourcing mentor.

In these meetings, the students described their project topics, their user research needs, and their planned sources. These meetings discussed crowdsourcing in general, how MTurk functions, and how the students wanted to use crowdsourcing. The projects independently developed the tasks and questions they wished to post on MTurk. The tasks were primarily exchanged via email, with occasional feedback. The researcher limited comments to clarifying confusing terminology, and providing format recommendations to help fit the survey template on MTurk.

The HITs (Human Intelligence Task - MTurk vernacular for tasks) were posted on MTurk by the researcher; this was a pragmatic choice to ensure the researchers had access to of the process and results. The students oversaw the posting of at least one HIT, where I explained each step and choice as it related to their task. The HIT results were downloaded and sent to the students, as they were completed. The researcher only rejected (did not pay) incomplete responses; additional quality judgments and analysis of the responses were left to the projects. After the projects completed all of the user research activities, they were asked to take a digital survey focused on their overall user research process and the strengths and weaknesses of the each method used. The survey was followed by a semi-structured interview that clarified answers but emphasized their experiences with crowdsourcing user research.

Data Collection

To answer the research questions this study primary focuses on participant responses to the end of project survey and follow-up interviews. Additional data was collected from a number of sources including: tasks and results on MTurk, observations, email exchanges, and project deliverables. These additional sources served as contextual support and a lens to help interpret the results of the survey and interviews.

The survey was setup on Google Docs and all student design project members were emailed a link. Between one and three students from each project responded to the survey. The survey covered project goals and information needs, but focused on the user research methods used, their perceptions (strengths and weaknesses) of the methods used, and how the results were applied to the project. The responses to the surveys were analyzed and consolidated to identify the way each method was used as well as its strengths and weaknesses. Each survey response was followed by an interview.

The interviews began by asking the students to draw their project process indicating where and what user research took place. This served both as a priming tool (as there was often a one to two week gap between survey and interview) and as a visual reference during the interview. The interview was divided into two sections, first I reviewed their survey responses, asking them to expound on their previous answers. The second half was focused specifically on four phases of their crowdsourcing experience: initial perceptions, use, results, and future use. The interviews followed a script, but took opportunities to explore relevant topics. The interviews were digitally recorded for later transcription and analysis.

Interview Analysis

The interviews were transcribed using three separate online crowdsourced transcription services (speechpad.com, transcriptionstar.com, and castingwords.com). Several checks were conducted to ensure the quality of each transcript. The overall quality of the transcripts was high, although some errors were found, most were related to topic jargon and colloquialisms, and were easily understood. The interviews were analyzed using qualitative content analysis (Elo & Kyngäs 2008). This approach is similar to grounded theory (Glaser & Strauss 1967) and is used to uncover emergent themes in the results without the use of predefined coding or categories that may preemptively constrain the results. The distinction is that qualitative content analysis leaves the emergent themes as data points and does not codify them into a theory. To conduct the qualitative content analysis each transcript was loaded into a spreadsheet with columns for line number, speaker, quote, and paraphrase. The researchers used this last column during analysis to paraphrase all quotes they felt contributed to the understanding of the research questions. These paraphrases would later be card sorted into themes.

To help establish validity of the transcript, three researchers familiar with this study independently analyzed the first three transcripts. All three researchers analyzed the first transcript. During this analysis they each paraphrased any quotes they thought contributed to understanding crowdsourcing as a user research tool. The three sets of paraphrasing were then compared and discussed by the researchers. Quotes with more than one paraphrase (identified by more than one researcher) were compared to see if corresponding interpretations had been reached through the independent reviews. Quotes with only one paraphrase

were examined to determine why the others had not commented. Two additional transcripts were similarly paraphrased by two researchers each and reviewed following the same process of comparison.

From these three interviews only a handful of mismatches were found. Where these occurred, they helped to clarify the paraphrasing. Quotes that were selected by only one researcher frequently had a matching paraphrase (from another researcher) attached to a neighboring quote. After the joint review of three interviews, key trends were emerging and only minor differences in paraphrasing were found. The lead researcher paraphrased the remaining eight interviews.

The 11 interviews contributed an average of 40 paraphrased quotes each. The quotes were grouped into themes using a card-sorting technique as shown in Figure 5.10. Each theme was individually analyzed and subdivided to identify common contributing factors, which are described in the results section below.



Figure 5.10 The card sorting process. Each slip of paper is color coded by participant and includes the original quote and the researcher paraphrase. They are shown in stacks representing different topics.

Results

The results for this study start with the fundamental data captured through MTurk to get an idea for the scope and quantity of activities involved. Next, the survey responses show how MTurk was incorporated. I then focus on the strengths and weaknesses of crowdsourcing as a user research tool. The qualitative content analysis of the interviews provides detail on the value and fit of crowdsourcing in a design project. A list of recommendations based on the participants' experience is included.

The data from MTurk (Table 5.3) shows the seven student design projects hosted a total of 26 individual HITs. Projects ran between one and four unique HITs each, but often ran iterations with minor changes to target a particular country or visual stimulus. In total, 810 responses were received; of these 98 were rejected as incomplete. Design students later reported discarding an additional 2% to 10% of the responses as unusable. Most of the HITs paid \$0.05 per response; although eight were paid a bonus due to added length or to attract attention when seeking target groups. The 98 incomplete responses highlight the diffi-

Table 5.3 An overview of all the 27 MTurk HITs posted by the projects during this study, showing the parameters, topics and number of questions asked by each tasks.

Project	Study	HIT/Format	Description	Reward (USD)	Responses Requested	Responses Received	Responses Rejected	Run Time (Days)	Avg Work Time (min:sec)	Avg pay (\$/hr)	Country
Furniture Concepts	1	Short Survey w/2 stimulus images	Brand awareness & impression	\$0.05	50	50	1	5	3:34	0.84	All
	2a			\$0.05	20	20	0	5	2:32	1.18	not India
	2b			\$0.05	20	20	0	5	2:38	1.14	not India
	2b			\$0.05	20	20	0	5	3:17	0.91	India
	2c	Medium > Multiple Choice	Impressions of work environment [picture] + Comfort	\$0.05	20	20	0	5	2:19	1.30	not India
	2d	Rating Survey w/Stimulus Image		\$0.05	20	20	0	5	3:30	0.86	not India
	2d			\$0.05	20	20	0	5	3:17	0.91	India
	2e			\$0.05	20	20	0	5	2:38	1.14	not India
	2e			\$0.05	20	20	0	5	3:13	0.93	India
AirPort Retail	1	Upload picture	Favorite shopping experience	\$0.05	50	50	0	5	na	na	All
	2	Upload pic + bonus		\$0.05	50	13	0	5	na	na	All
	3	Short - survey	Your favorite shop & why	\$0.05	50	50	1	5	2:06	1.43	All
	4	Link to short web survey		\$0.10	30	20	1	5	2:19	2.59	not India
In-Flight Experience	1a	Long - Mixed Survey	Flying Preferences +	\$0.10	40	40	2	5	9:52	0.61	USA
	1b		Impressions of other Cultures	\$0.05	40	40	1	5	12:51	0.23	India
Car Sharing	1a	Long - Quantitative Survey	Transportation habits +	\$0.07	50	50	0	5	4:21	0.97	not India
	1b		Impression of car sharing	\$0.10	50	50	1	5	3:11	1.89	NL
	2	Long - Quantitative Survey	Transportation solutions	\$0.25	20	6	0	5	2:31	5.96	NL
Insurance Policy	1a	Short - Mixed Survey	Love/hate about insurance company	\$0.10	20	1	0	7	2:19	2.59	NL
	1b			\$0.05	20	20	1	7	4:17	0.70	not India
	2a	Short - Open Question Survey	What makes you curious	\$0.05	20	2	0	7	3:25	0.88	NL
	2b			\$0.05	20	20	0	7	3:56	0.76	not India
eHealth	1	Medium - Mixed Survey	Computer habits	\$0.07	40	40	0	5	4:31	0.93	not India
Neighbors & Community	1	Short - Open Question Survey	Describe your neighbors	\$0.05	50	50	0	5	3:40	0.82	not India
	2a	Short - Open Question Survey	Discovering the good in people	\$0.05	24	24	0	5	6:00	0.50	All
	2b			\$0.05	26	26	1	5	5:28	0.55	not India

culty in gathering responses to tasks that deviate from standard practice (upload a picture and send a message for a bonus) or that highly constrain geography (Netherlands only). The overall success of these HITs (as perceived by the design projects) demonstrates the diversity of what is possible, in terms of both topic and format of the tasks.

Survey Results

In addition to HITs on MTurk, each student design project used two to five additional research methods (Table 5.4). These methods include expert interviews, literature study (journals and books), interviews, focus groups, contextmapping (Sleeswijk Visser, 2009), observation, and field studies. While not reported as a formal method, all projects spent some time gathering background information online. These methods served as both information sources for their projects and a point of comparison for judging crowdsourcing as a user research tool.

Table 5.4 An overview of the various user research methods used by each project as reported in the follow-up surveys. Each project used between three and six different methods.

Project Topics	CS MTurk	Web Search	Interview	Focus Group	Context Mapping	Literature	Survey	Field Study
Furniture Concepts	✓	✓	✓		✓			✓
Airport Retail	✓	✓	✓	✓		✓	✓	✓
In-Flight Experience	✓	✓	✓			✓		
Car-Sharing Service	✓	✓		✓	✓		✓	
Insurance Policy	✓	✓		✓			✓	
eHealth	✓	✓	✓		✓			
Neighbors	✓	✓	✓		✓	✓		✓

The relationship between the methods employed is not readily apparent in Table 5.4. The students described substantial diversity in each project’s employment of different methods. Some projects selected specific methods to target different information needs, others applied the methods in complementary fashion to gain multiple viewpoints, while others simply replicate questions with multiple methods to increase participants. While there is significant diversity in topics, information needs, and application of crowdsourcing, the common thread among the design projects is their use of crowdsourcing as an early exploration tool, pretesting for later activities, and adding diversity or quantity (including users who are typically out of reach of traditional methods).

Going beyond implementation, the research survey focused on the perceived strengths and weaknesses of each method based on each project’s experiences. Table 5.5 lists the responses specific to crowdsourcing. Commonly expressed benefits emphasized the ability to obtain a lot

Table 5.5 The strengths and weaknesses of crowdsourcing as a user research tool, as reported by students each project in the surveys. They appreciated the speed, reach, diversity, and insights they were able to receive, but did not like giving up control of who answers, filtering out unusable responses, or the inability to clarify responses.

Project Topics	Strengths	Weaknesses
Furniture Concepts for Future Work Environments	Short & fast, results quickly, diverse & International participants, surprised by results [the crowds awareness of a small brand]	Responses less rich [in contrast to interviews and context mapping], not every participant is serious about responses
Airport Retail - Premium Products for Women	A lot of entries allow you to generalize insights, confirm findings, down select from suggestions in literature, considerably quicker, many people, international [responses]	Some entries need interpretation, little to no quality control when respondents not socially bound, hard to steer, misuse
In-Flight Experience for Chinese Passengers	Ability to reach outside known associates, many responses	You must ask the questions properly or you will get answers that are lacking
Market Entry Plan for Car-Sharing Service	Lots of information, no waiting, no recruiting, quickest and easiest way to collect [user] data for statistical analysis, supported findings from context mapping	Challenging to target the 'right' users [drivers w/o cars within a small geographic region], need to spend time to properly develop a good questionnaire and evaluate responses.
Insurance Policy Design, for GenerationCY	Data from people not on your list (inspiration or point of comparison), ability to use media, no search for participants, answers wellCthought (not on the spot like interviews), participant freedom to answer	Can't ask follow-up questions, no giving direction/clarification of questions, cannot empathize
eHealth for Digital Natives	Quite informative	Initially struggled to interpret [qualitative responses]
Stimulate Contact between Neighbors	A lot, quickly, structured results	No control over who answers [can't hand select]

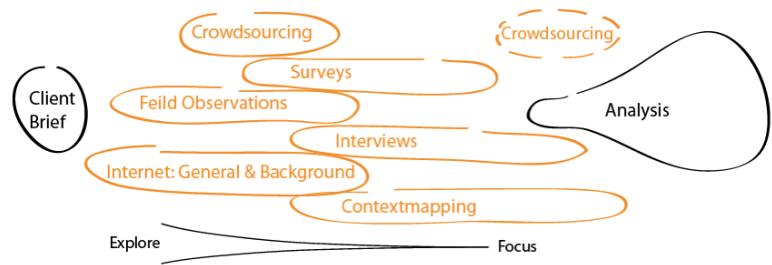
of results quickly from a large, diverse, and international population. MTurk was seen as a valuable addition to other methods, as confirmation, and an outside (non-target user) perspective. There was less agreement regarding reported weaknesses are, but the student design projects highlighted the lack of control in selecting participants, less depth or richness in the responses, difficulties in developing “good” questions, and instances of unusable responses. The interview analysis below provides additional insight into these and other limitations.

Interview Results

Since these projects were tied to course requirements they all followed a similar design process (Client Brief, Vision, User research, Analysis, Concept, Development, and Materialization). The process diagrams drawn at the start of each interview confirmed this consistency. The diagrams also showed relative consistency in the timing of each method and its duration, revealing the fit of crowdsourcing in the user research process. Figure 5.11 depicts the methods used during the research process along with relative placement, overlap, and duration.

It shows crowdsourcing early in the process though usually after some initial Internet research and field observations when possible. These initial observations were usually followed by either interviews or contextmapping. A couple of groups chose to re-employ crowdsourcing for additional inputs or to confirm findings from other methods. Informal analysis was often conducted during the research phase and later formalized at the conclusion of user research.

Figure 5.11 An averaged alignment of the user research employed by the student design project. While each project did not use all methods shown the relative timing and duration of the tools used was consistent (consolidated from interview diagrams).



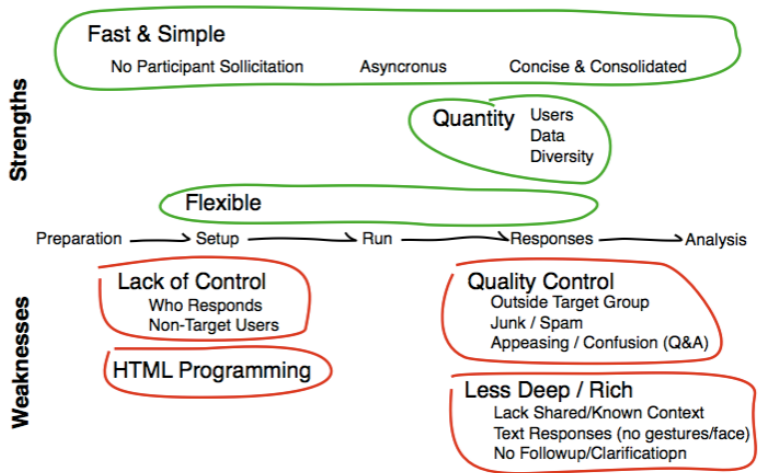
Not surprisingly, the interviews largely echoed the survey results, but provided depth and additional insights into the value and fit of crowdsourcing as a user research tool. The interviews expound on the strengths and weakness student designers experienced while using crowdsourcing to inform their projects. The transcript analysis revealed a number of recommendations that served as a refresher on survey design as well as highlighting some important considerations when gathering user inputs online.

The value of crowdsourcing as a user research tool is observed in the descriptions of how the results contributed to each project. The transcript analysis revealed how the projects appreciated crowdsourcing's fit into their research as groundwork to formulate questions for study by other methods, as an additional or expanded source, to gain a broader or external perspective, and to confirm or "down-select" results gathered through other methods. Each project was able to accomplish specific information goals and incorporate the users' insights into their projects. The Furniture project was able to connect user impressions of the brand to the desired corporate image. The Airport Retail project looked at user experience in shopping to down-select from common retail design elements. The crowdsourcing results provided the Car Sharing group "statistical proof" for the insights gained through interviews. This was significant because the corporate sponsors were accustomed to numeric market data. Crowdsourcing became the primary source of information and inspiration for the Insurance Policies

project. The anonymity allowed the Neighbors project to gain more candid responses to personal questions about prejudices toward people in the respondents' community.

While these insights demonstrate the value and fit of crowdsourcing within each project, the interviews also provide detail into specific strengths and weaknesses. The additional detail also highlights contributing factors and connects them to different phases of the user research process. Figure 5.12 maps the strengths and weaknesses on a representative timeline showing where and how they impacted the user research process, and calling out some of the contributing factors that support the benefits and concern

Figure 5.12 Strengths and weaknesses of crowdsourcing as a user research tool, as reported in interviews, mapped along a notional timeline of the user research process. Benefits include the speed, ease, and flexibility of crowdsourcing as well as the abundance of responses. Weaknesses focus on a lack of control in who responds and the variable quality of their responses, which also tend to be less rich.



The benefits emphasize the speed, simplicity, and flexibility of crowdsourcing as well as the quantity of responses. The speed and low effort are primarily a result of eliminating participant solicitation and adding asynchronous or automated data collection. The relative ease in analyzing results comes from concise responses and structured results. Because of the relative ease of the process, three of the projects conducted less formal analysis and placed less value in the results. Receiving a high volume of responses from a large and diverse group of users provided a much broader picture by gathering insights from users who are generally out of reach of traditional methods. The flexibility of the MTurk application allowed the student design projects to gather the information they desired by deploying a wide variety of tasks to fit their projects' needs.

The primary weaknesses are largely connected to a general lack of trust, an inability to hand-select and meet participants, and navigating an unfamiliar and untested tool. The lack of control in selecting participants evoked concerns over reliability and representativeness of the results. For some groups (despite preconceptions) the analysis of the results alleviated these concerns and revealed unexpected value in providing a broader perspective. In addition, the results uncovered similarities (and differences) in responses from people outside the target group. While a benefit to some projects, for other projects the diversity of respondents caused difficulty or confusion interpreting the results.

In addition to the desire to control who responds, there was a concern regarding useless or spam responses. In the end, less than 10% of the responses were considered spam, and they found the spam easy to identify and discard. The presence of spam, however, induced concerns of trust in the cumulative dataset (they asked: How do you know the others are legitimate?). Interestingly, two projects noted that they had a similar problem with interview participants providing appealing responses.

Despite the reported appreciation for the crowdsourcing results, the student designers consistently perceived them as providing information that was less deep or rich than the information gathered from methods where they are face-to-face with participants. The in-person methods benefit from a shared experience or context, knowing or meeting the person, and nonverbal cues. In addition, the ability to clarify questions and answers reduces confusion or ambiguity during analysis. In contrast, two of the projects thought the large number of diverse responses provided depth and richness, though it was seen as an addition not a replacement for the richness gained through interaction with users.

Three participants with no programming experience expressed apprehension over HTML programming. While programming is not strictly necessary on MTurk, it is definitely beneficial (for customizing HITs and formatting output). This would be a significant barrier for people uncomfortable with even basic programming.

Recommendations for Use

Based on their experiences with this study, the student design projects provided a number of insights for success when using crowdsourcing as a user research tool. Those familiar with guidelines for survey design or general crowdsourcing will not be surprised. The recommendations

highlight elements of the process that are especially beneficial at this unique intersection (with examples, and how they support crowdsourcing for user research).

- Write clear instructions and questions (avoids confusion, helps them think about topic, is easier to answer properly, improves response quality, simplifies analysis)
- Combine different types of questions (short/quantitative for fast analysis, open for qualitative context and quotes)
- Include demographic questions that help interpretation (nationality or location)
- Be aware of your audience (cultural differences)
- Let people share, simple open-ended questions (What makes you curious?)
- Avoid complex questions that can be confusing and hard to answer (Describe several factors of a pleasurable shopping experience? Vs. Why do you love shopping?)
- Allow people to complain about unpleasant topics (Why do you hate your insurance company?)
- Provide other (and null) response options
- When asking for pictures, provide space for description (ask why/ describe)
- Use a stimulus image to encourage responses
- Improve and repost questions (learn from your mistakes)
- Run multiple HITs (parallel or sequential) as needed

In short, be clear and ask an interesting mix of questions to encourage people to share personal insights.

Discussion

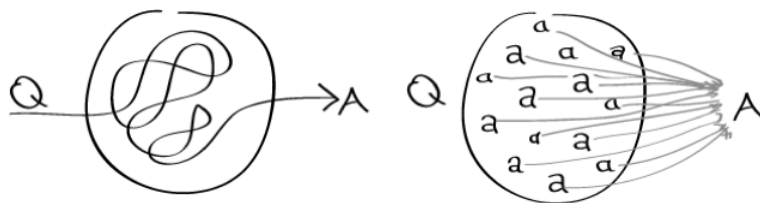
This study establishes that crowdsourcing can be a valuable tool for user research. The seven student design projects primarily employed crowdsourcing as an early exploration and survey tool. The inherent speed, flexibility, and ease in reaching a large and diverse population allowed them to gather a lot of information from people who were typically beyond their reach.

The flexibility of MTurk is reinforced by the wide variety of formats and topics hosted by the projects: from simple demographics and rating scales, to describing preferences and reflections on personal experiences. Also included were three HITs that asked participants to upload an image or links to images. The responses served as an early foundation

to broaden each project's understanding of the project context, as well as enriching, expanding, or confirming insights gained through other sources. The large, concise, and structured results made them easy to absorb and incorporate into the project.

While there are several benefits of this approach, it is not without drawbacks. These include a lack of trust, both in who participates and the quality of their responses; lack of control over selecting participants (akin to coordinating interview participants); and not recognizing the potential benefits of participant self-selection. Additionally, the lack of contact with users diminishes the perceived depth and richness of the information that is gathered. These barriers run contrary to a research goal that aims to build a more complete understanding of the user to inform the design process. While some student design projects posed questions that gathered deeper responses and others found richness in the diversity of responses (Figure 5.13), crowdsourcing did not replace the emotional connection and contextual insights gained through meeting and observing users in person.

Figure 5.13 A visual depiction of gaining rich information from users. The left shows a question that is answered by traveling a single path that winds through a specified design space in order to reach a richness of information, desired in the answer. In contrast the diagram on the right depicts a crowdsourcing paradigm where richness can be gained by collecting numerous inputs throughout the design space.



Some issues identified by the student designers have more to do with inexperience than with crowdsourcing. This was observed in their struggles to design questions and to analyze responses. Receiving the crowdsourcing results built an understanding of how people respond to different types of questions. This insight led to many comments on the importance of good and clear question design. The students learned that the quality of questions greatly influences the quality of the answers. To ensure you receive the type of information you need, you should develop questions that encourage people to share their insights into your topic.

While half of these results may not surprise the crowdsourcing community, and the other half may not surprise seasoned user researchers, it is the intersection of these two fields that provides new insights. These insights reveal crowdsourcing as a fast and powerful tool to be added to the arsenal of user research tools.

Conclusions for Framework

The value of crowdsourcing as a user research tool comes from the speed and ease with which designers can reach a large diverse population that includes people willing to answer questions and provide information. The crowd adds significant breadth and quantity to user research results. The flexibility of the MTurk application supports a wide variety of tasks and topics making it possible to gather information in multiple methods or formats. These characteristics of a crowdsourcing application remove the need to find participants and also automate data collection, making it a fast and flexible addition to the user research process. This newfound access to users fundamentally changes the user research paradigm providing an option to access users early and often throughout the design process. Rather than spending significant time and energy digging for the rich insights of a few select users, designers can ask a huge audience of potential participants, making it possible engage users in an iterative cycle as knowledge grows and the design evolves. The richness of the insights no longer comes from how deeply a designer explores a topic with a few people, but from the amalgamation of insights from a much broader and diverse pool of participants, both inside and outside the target user population.

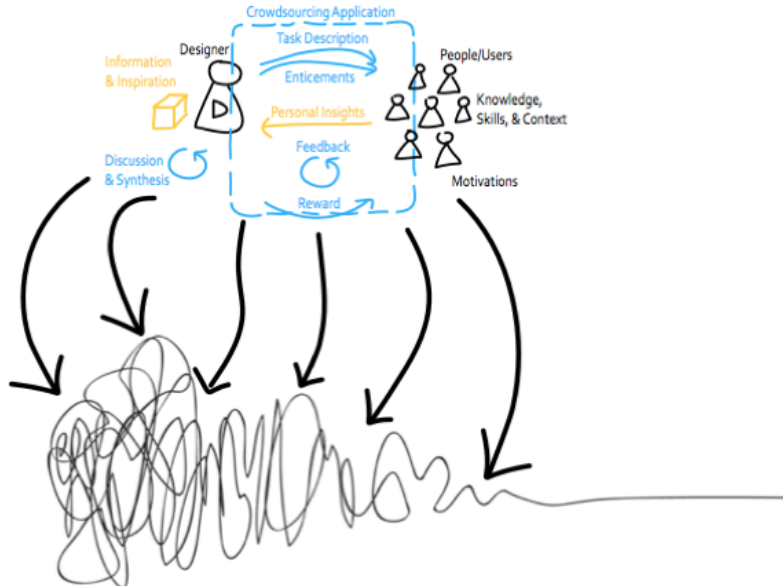


Figure 5.14
Crowdsourcing as a user research tool provides a fast and simple way to access a lot of diverse participants, providing the opportunity to iteratively engage users and provide timely user insights throughout the design process with minimal cost and effort.

Study 4: Tips for Design Practice

Tip 1: Be clear when developing tasks.

Tip 2: Embrace the opportunity to reach outside your target population to compare and contrast and develop a broader perspective.

Tip 3: Depth can be found in letting people share (simple open questions).

Tip 4: Richness can be found in the amalgamation of many diverse responses.

Tip 5: Post early and often throughout the design process.

Tip 6: See above for recommendations for use as identified by the study participants.

5.3 - Study #5: Fast and Good

Designers' Skepticism: "So it's faster than most user research, but how fast? And won't the increased speed ruin quality?"

This study examines the speed of crowdsourcing for gathering user inputs. The preceding studies have gathered user information faster than many user research methods by eliminating the need to identify and solicit participants, and by automating data collection. While results from the preceding studies have consistently received the desired number of responses within a week, these speeds are conservative in comparison to other crowdsourcing activities. This study examines the speed of six user research tasks specifically designed for speed. I will explore the upper limits of speed for gathering user information with the goal of increasing access to user information in the initial phases of the design process.

Research Questions

- How fast is crowdsourcing for user research tasks?
- *What are the implications to bring user inputs into the early stages of the design process?*
- *Does the emphasis on speed negatively impact response quality?*

Literature has a number of examples demonstrating the speed of crowdsourcing with response rates as fast as 50 to 170 submissions per

hour, but these are generally “click-work” tasks requiring little effort. Some of these studies manipulate the reward to measure the effect on both speed and quality (Borst, 2010; Feng, Besana & Zajac, 2009). The results show how higher pay produces faster results, but excess reward decreases quality. The decrease in

This study is based on the following publication:

Tidball, B., Mulder, I. and Stappers, P.J. (2017). Crowdsourcing As An On-demand User Research Tool. Submitted to *ACM Transactions on Computer-Human Interactions*.

quality is a result of attracting spam responses from people trying to game the system, because the possibility of a reward outweighs the potential of being rejected for unacceptable responses. In addition

DiPalantino and Vojnovic (2009), show how word choice within the task description can influence participation. Schwarz (1999) shows how the questions shape the answers, recommending clear and concise phrasing. These findings have led to recommendations in the “how to” guidance on many crowdsourcing applications.

Figure 5.15 This study focuses on the speed of responses. Based on what literature and the previous studies illuminate about the crowd workers and their motivations, that knowledge can be applied to the Task Description, Enticements, and Reward to entice much faster responses.



Context

This study was conducted in conjunction with two workshops within the Industrial Design Engineering faculty at Delft University of Technology. The workshops were designed as an introduction to crowdsourcing as a method for gathering inputs from users to inform design. In order to demonstrate the process, live crowdsourcing activities were conducted during each workshop. The live demonstrations would need to generate results quickly (within the 2-3 hour workshops) driving an emphasis on the aspects of task design that would accelerate response time.

The first workshop was an optional element in a first year design course called Exploring Interactions for students of the Design for Interaction Master (DFI). In this course students are introduced to user research and complete an individual design project. This workshop had two sessions, with 13 and 16 first year design students respectively. Due to the time constraints of the workshop and the students' limited experience with user research, the crowdsourcing tasks and questions were designed in advance and without their input. The tasks were setup, built, posted, and run as a demonstration during the session with the incoming responses displayed live. While waiting for the responses, the students participated in a paper crowdsourcing exercise. The exercise

introduced some fundamentals of survey writing followed by the group participating as a crowd in three of the paper crafted surveys. At the end of the sessions, the crowdsourcing responses (received during the first hour) were examined with discussion on their quality as perceived by the students.



Figure 5.16 Students participating in the Exploring Interaction workshop, participating as crowd-workers, selecting and responding to their choice of available tasks.

The second workshop had one session with nine students, divided into two teams. This workshop was one of several options offered as part of “DFI Master Identity Day,” a semester kickoff event hosted by the industrial design faculty. The teams were a mix of design students from each year (undergrad through masters). During the three-hour workshop the student teams had to create their own crowdsourcing tasks (with guidance on crowdsourcing and hosting support) and use the results to develop and present a poster. Figure 5.17 shows the posters created by the students to summarize the user insights they gained from the responses to their crowdsourcing tasks.



Figure 5.17 Students presenting the processed results they received through crowdsourcing tasks during the “DFI Thursday” workshop.



Method

Each of the crowdsourcing tasks conducted for this study used Amazon’s MTurk. The workshops took place in early afternoon central European time, and tasks were constrained to Turkers registered in India (the largest population of Turkers available at that time of day).

For the crowdsourcing workshop for the Exploring Interactions students, the tasks were created in advance without input from the students. Special attention was given to developing tasks that were likely to generate responses within minutes (not days). To accomplish this, the list below identifies several recommendations (along with their intended benefit) from literature, MTurk's best practices guide, and previous experience were applied:

- Broad topic (wide audience)
- Slight overpayment (attract attention, but limit spam)
- Clear and concise title and description (set expectations of effort)
- Clear and concise questions (limit effort and avoid ambiguity)
- Short task with common/familiar structure (limit effort)
- Ask questions that allow people to share experiences or opinions (enjoyable)
- Entire task (including submit button) visible in task preview (known effort)
- Appropriate answer box size (indicate appropriate length of expected answers)
- Short task time (indicates limited time/effort to complete)
- Short run time (indicates limited task availability, now or never)

Two generic topics – vacations and gadgets – were chosen for these tasks to allow a broad audience to respond. The tasks were built in MTurk, starting with the standard survey template (familiar format for participants on MTurk), which includes basic demographic questions. The task properties were set as prompted by MTurk with an emphasis on clarity. Figure 5.18 shows a screenshot of the settings for reward, number of assignments, time allotted, expiration, and auto approval time for the HIT.

Gender and age questions from the template were kept but the other template questions were deleted. Three topic questions were added including two short answer questions with a representatively small space to enter their answer and one longer question with an answer box appropriate for a one-sentence description (see Figure 5.19).

At the end of the workshop, the responses were shown to the students who then discussed the various quality and utility of the responses as well as an expectation of how users respond to survey tasks.

Describe your HIT to Workers

Title
 Answer an survey about your opinions
Describe the task to Workers. Be as specific as possible, e.g. "answer a survey about movies", instead of "short survey", so Workers know what to expect.

Description
 Give us your opinion about our products
Give more detail about this task. This gives Workers a bit more information before they decide to view your HIT.

Keywords
 survey, demographics
Provide keywords that will help Workers search for your HITs.

☐ This project may contain potentially explicit or offensive content, for example, nudity. [\(See details\)](#)

Setting up your HIT

Reward per assignment
 \$ 0.05
Tip: Consider how long it will take a Worker to complete each task. A 30 second task that pays \$0.05 is a \$0.00 hourly wage.

Number of assignments per HIT
 50
How many unique Workers do you want to work on each HIT?

Time allotted per assignment
 15 Minutes
Maximum time a Worker has to work on a single task. Be generous so that Workers are not rushed.

HIT expires in
 1 Days
Maximum time your HIT will be available to Workers on Mechanical Turk.

Results are automatically approved in
 2 Hours
After this time, all unreviewed work is approved and Workers are paid.

[Advanced >](#)

Figure 5.18 A screen shot of the settings used on MTurk (parameters for all tasks shown in Table 5.5). The simple text, attractive pay, and short durations were all selected to encourage participation and indicate low effort.

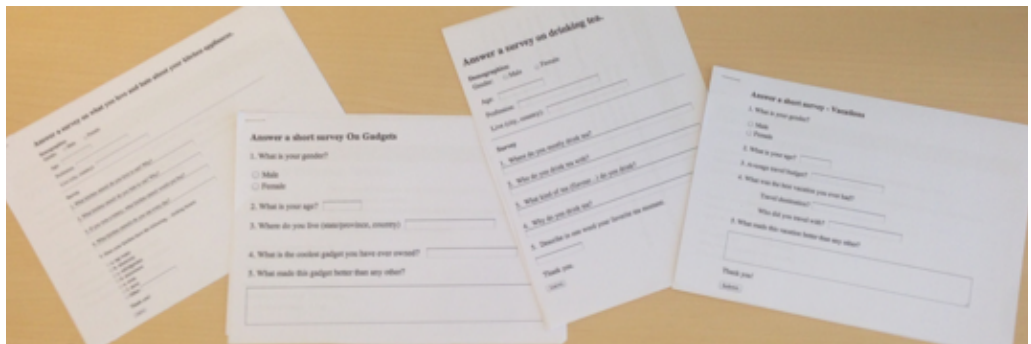


Figure 5.19 Printed copies of the survey tasks as they appeared to users on MTurk for both the Exploring Interactions and DFI Thursday workshops. They included basic demographic questions followed by two to five user research questions.

For the DFI Thursday workshop the students were asked to identify their own topic for the crowdsourcing task. They developed five questions that would be added to four demographic questions preselected from the MTurk survey template. They were given a brief overview of best practices for creating a HIT with recommendations geared toward speedy responses. To aid in the process (and ensure data capture for this research), I posted the surveys with an existing MTurk account. The parameters and settings for the tasks were set identically to the previous workshop (Table 5.5).

Once the survey tasks were posted a couple of the students asked if they could create a second task requesting pictures, as they had seen results from one of my of previous studies. For the picture tasks the “file

Table 5.5 The setup parameters for each of the six crowdsourcing tasks.

Title	Vacation	Gadgets	Drinking Tea	Kitchen	Tea Pic	Kitchen Pic
Date	6-Oct-11	10-Oct-11	9-Feb-12	9-Feb-12	9-Feb-12	9-Feb-12
Qualification Requirement	HIT Approval Rate > 95%, Location is India				none (not an option)	
Number of Assignments	50	50	50	50	20	20
Reward	\$0.05	\$0.05	\$0.05	\$0.05	\$0.10	\$0.10
HIT expires in	1 Day	1 Day	1 Day	1 Day	1 Day	1 Day
Assignment Duration	15 Minutes	15 Minutes	10 Minutes	10 Minutes	10 Minutes	10 Minutes
Auto Approval Delay	1 Day	2 Days	2 Days	2 Days	na	na
Task Design	5 Questions	5 Questions	9 Questions	9 Questions	1 Picture	1 Picture

upload” template was used with a single statement of “Please submit a picture of ____.” Previous experiences with picture tasks had returned slow response rates so the reward was doubled to \$0.10 per picture and only asked for 20 responses, with all other settings remaining the same.

As the responses were received they were given to the student groups to be analyzed and incorporated into a poster and presentation.

Data Capture and Analysis

Timing data was collected automatically by MTurk and is part of the standard results spreadsheet (*.csv file). The “file upload” template returns fewer data columns and some of the results had to be calculated separately.

Responses were judged by the author on a three-point scale of useful, ambiguous, or reject. The simplicity of these tasks made judgment straightforward. The most obvious junk was rejected and anything questionable was labeled ambiguous (see results below).

Results

In response to the six HITs in this study, 240 responses were received. Eleven were discarded, 10 were identified as ambiguous, and 11 pictures were redundant. This left a total of 202 useful responses at a cost of \$15.40. Table 5.6 shows a breakdown of the results for each study. The average response rate across all six studies was 13.8 responses per hour, or 11.6 usable responses per hour, with the first hour being the most productive with an average of 19.2 responses. The picture tasks received a slower response despite the higher reward.

To visualize these responses, Figure 5.20 shows the cumulative number of responses received over time. Not surprisingly the response rates for

Table 5.6 Summary of the results for all six crowdsourcing activities, including time data and quality assessments.

Title	Vacation	Gadgets	Drinking Tea	Kitchen	Tea Pic	Kitchen Pic	Average
Avg Time Per Response:	2 min 19 sec	2 min 25 sec	2 min 0 sec	3 min 25 sec	1 min 39 sec	1 min 40 sec	2 min 15 sec
Number of Assignments Request/Receive/Reject:	50/50/0	50/50/0	50/50/1	50/50/0	20/20/11	20/20/11	40/40/3.5
Effective Hourly Rate:	\$1.30	\$1.24	\$1.50	\$0.88	\$3.63	\$3.61	\$2.03
Time to Complete All Tasks:	5 hr 11 min	3 hr 57 min	2 hr 8 min	2 hr 24 min	2 hr 21 min	2 hr 37 min	3 hr 6 min
Response Rate (per hr):	9.6	12.7	23.4	20.8	8.7	7.6	13.8
Usable Responses:	46	45	49	41	9	12	202
Reject Responses:	2	3	1	3	0	2	11
Ambiguous Responses:	2	2	0	6	5	6	21
Redundant Responses:	-	-	-	-	6	-	6

each pair of tasks show similar trend lines suggesting that the type of task and when it is posted will influence the speed of responses. Next Figure 5.21 shows the distribution of usable responses per task. As can be seen, the question-and-answer tasks had few unusable (rejected) responses, while the picture submissions had far more ambiguous responses with questionable utility.

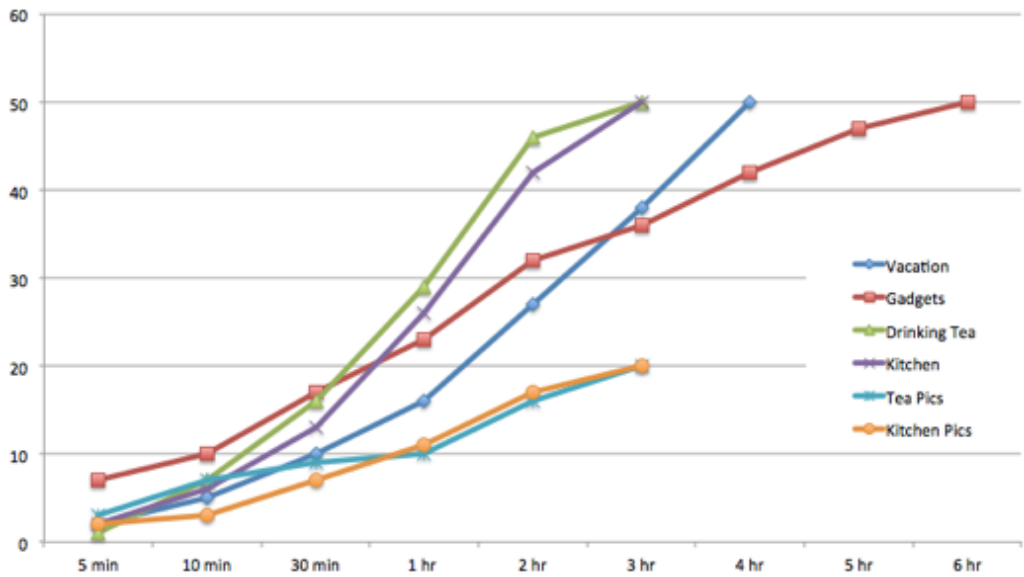
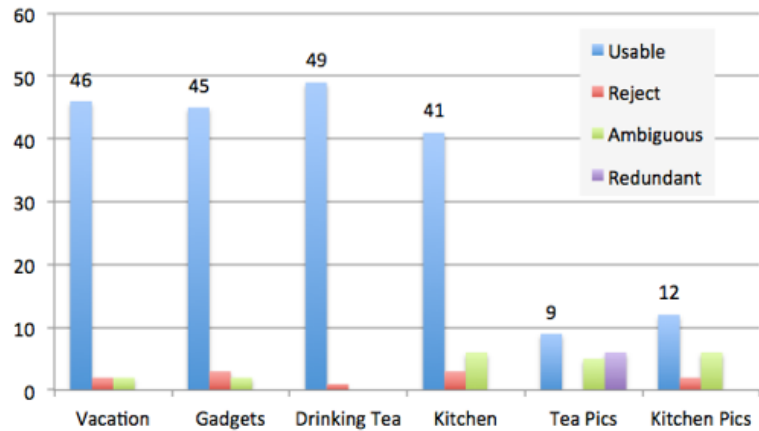


Figure 5.20 The cumulative number of responses received over time for all six studies (note: the time access is not linear, on a linear scale the response curves start off with high response rates and then slowdown).

Overall, 84.2% of the responses were judged to be useful, meaning they provided coherent answers to the questions asked. Looking solely at the survey responses, 90.5% (181/200) were judged to be useful. On the other hand, the picture tasks had substantially fewer usable pictures 52.5% (21 of 40), where two were rejected, 11 were ambiguous, and

Figure 5.21 The judged quality of responses to all six tasks. The text responses were mostly usable while the picture submissions were far more ambiguous in their utility.



six were redundant (Figure 5.21). Despite the high number of irrelevant pictures, the students found the remaining images helpful for their understanding and they provided a visual context for the survey responses.

Discussion

The results above demonstrate how crowdsourcing can be used as a near on-demand user research tool. Through conscientious design of the task we experienced response times that averaged 5 responses in 10 minutes. While not instantaneous, it is fast enough to receive user inputs during the course of a meeting. This speed provides a means for users to inform the early stages of the design process, ensuring that early project decisions are user centered. Early user input can help inform foundational strategic decisions that often set the direction of the design project. Additionally, the emphasis on speed did not have a negative effect on the quality of the responses. During the synthesis of the responses student designers were able to easily identify and discard spam and low/no value responses. The unused responses accounted for roughly 15% of all responses; in contrast, we discarded 10% of responses generated though our other research activities on MTurk. Due to the increased payment, we expected significantly more spam responses. But the results indicate that the same considerations that help encourage faster responses (clear tasks that require minimal effort) also allow or encourage users to provide clear and useful responses. The higher rate of spam for picture responses is likely because the task is quicker (than the survey) and more difficult for the requester to justify rejecting a response.

Conclusions for the Framework

This study was able to show how crowdsourcing as a user research tool can be greatly accelerated through conscientious task design. To en-

Figure 5.22 By focusing on creating user research tasks paired with enticements for fast results, the tasks averaged more than 13 responses per hour without negative impacts to quality, though the response rates did not reach the speed of click-work. For user research, this provides an opportunity to gather user insights anytime questions arise throughout the design process.

courage fast responses, tasks should seek a wide audience appeal with clear wording, ample reward, and minimized time and effort indicators. As highlighted in Figure 5.22, careful attention given to these parameters in the design of the task description and the corresponding enticements also appears to encourage quality responses (average 84.2% usable). In other words, the careful attention given to designing a faster task by clearly stating what is desired makes it easier for responses to be accomplished, improving both their speed and quality.



With some response rates approaching 20 responses per hour, crowdsourcing can be used to gather inputs from users as questions arise throughout the design process. This enables designers to bring user insights into the initial and conceptual stages of the process to help influence critical strategic level decisions with user provided information. This early and flexible opportunity can aid projects in starting off on the right foot and can answer questions as they arise without the typical delay incurred for most user research methods.

While the speed demonstrated here was aided by the simplicity of the tasks and topics, careful attention to the design of tasks can improve the results of any crowdsourcing activity. Longer, difficult, or nonstandard tasks will decrease speed. This was demonstrated by the picture upload tasks which, despite having double the reward, had a significantly slower response rate of eight responses per hour. Attention to task speed actually taps into a number of motivational elements that benefit both speed and quality, including the simplicity and clarity of the title, description, and questions.

Study 5: Tips for Design Practice

Tip 1: To encourage fast responses, design tasks for a wide audience with clear tasks, ample reward, and minimize time and effort indicators.

Tip 2: Set task parameters to reflect a fast and simple task. Be sure the task matches the settings.

Tip 3: Take time to design clear tasks. Clear simple questions result in clear simple answers, and are easier to sort and analyze. Overly specific or detailed questions add confusion.

Tip 4: When deeper more thoughtful responses are desired, use simple priming questions to initiate thinking, or use open questions that allow people to express themselves freely. These responses will be broader and more ambiguous requiring additional analysis.

Tip 5: Ask for pictures especially for unfamiliar contexts or cultures, but expect a slower response rate and more spam. (You can repost pictures as stimulus images).

5.4 - Study #6: Finding Target Users

Designers' Skepticism: "How do I get responses from a specific user group, not a crowd"

This study examines the use of qualification tasks as an approach to reach and identify target users in crowdsourcing activities (Figure 5.23). Once experts or target users are identified they are asked to complete a second task that seeks expert-user inputs. Through the previous studies the topics of the crowdsourcing tasks were largely open or generic, meaning most people should be able to provide input, and demographic characteristics were used to parse the responses as needed for comparison. However, there are times when inputs are needed from a highly constrained user group.

Figure 5.23 This study examines the use of qualification tasks to identify experts or target users. Once the experts are identified they are asked to complete a second task that addresses the specific expert-user inputs to determine if highly specific expertise can be reached from within the crowd.



The expense of engaging people in common user studies often necessitates that researchers find and filter participants in advance. In crowdsourcing this paradigm is upended, shifting the role of selecting who participates to the participants themselves. Even with this role rever-

sal, it is still valuable and often necessary to know who is responding and whether they possess the necessary expertise from the researchers' perspective.

Research Questions

- How can crowdsourcing reach people with specific expertise or knowledge?
- *What is the role of qualification tasks in identifying specific expertise?*

This study is based on the following publication:

Tidball, B., Mulder, I. and Stappers, P.J. (2017).
Crowdsourcing As An On-demand User
Research Tool. Submitted to *ACM Transactions
on Computer-Human Interactions*.

In the how to guidelines on MTurk, and in literature (Alonso, Rose, et al. 2008; Kittur, Chi, & Suh 2008) it is possible to create “qualification tasks” as training or filtering tools to create a pool of qualified workers for future tasks. These qualification tasks are employed as a

barrier for entry, where the training helps ensure subsequent tasks are completed properly. The intent is that future work is of higher quality. For conducting user research, the qualification task may be an effective method to seek out and identify people who have the desired expertise and represent a target user group.

Context

This research was conducted in cooperation with two pre-existing user-centered design projects. The projects were selected because they had user groups that were constrained and difficult to reach, but for different reasons. The researchers for these projects were both doctoral candidates in industrial design, who had conducted several iterations of user research, design, and prototype testing. As with several of the previous studies, data collection was accomplished through HITs posted on MTurk.

The first topic was designing for children with autism spectrum disorder. Designing for these children is difficult because their ability to interact socially is highly constrained and they are not always able to answer questions. Instead designers need to engage caregivers (parents, teachers, therapists, etc.) who live, teach, and play with these children. These caregivers are a small subset of the general public.

The second topic was on social connectedness and how to design products so people can feel and be more connected to loved ones who live in different places. While most people want to stay connected with friends and family, it is difficult to find and identify people who understand and

can communicate the emotional, technological, and generational differences involved with helping people feel more connected.

Method

This study was designed with three phases, while running the two topics (autism and connectedness) in parallel. The first phase was a qualification task designed to determine the expertise of respondents as it related to the respective topic. This was followed by a second task designed to gather user inputs. The second user research tasks were first available to the respondents of the qualification phase. For the third phase the user research tasks were made available to the public. The responses to all follow-up tasks were also rated on expertise to allow for comparison.

Tasks

The qualification tasks were designed by the topic experts with some guidance from me to ensure fit with the crowdsourcing application. The goal of the qualification tasks was to allow the researchers to assess if respondents possessed the appropriate expertise to be included in future studies. Figure 5.24 shows a screenshot of a qualification task. The qualifications included basic demographics and open-ended questions to help reveal respondents' experience on the topic not a self-report of their perceived knowledge or abilities.

The two qualifying tasks were posted on MTurk, seeking 150 responses per topic and offering \$0.05 per response. The number of responses for each task was divided evenly between the US, India (the two primary groups on MTurk), and everywhere else (50/50/50). This division was done to receive a more even distribution of responses and help identify cultural differences in the results.

The second set of tasks was adapted from research previously conducted by the topic experts. The tasks were similar in effort and easily adaptable to an online survey format (see Figure 5.25). The task for the autism study consisted of rating six lunch box designs aimed at supporting children with autism. The social connectedness task was split between asking about the important aspects of communication related to feelings, technology, and limitations and then asking them to evaluate a design concept aimed to support passive communication.

Rating Responses

The responses were given to the respective researcher for evaluation

Answer a short survey about Children with Autism Spectrum Disorder.

1. What is your connection to children with Autism Spectrum Disorder? (parent, therapist, teacher, doctor, etc)
2. What is their diagnosis?
3. What is their age?
4. Are they verbal?
5. What do you see as the biggest challenges these children strive to overcome?
6. What products would you recommend to help them overcome these challenges? How do they help?
- Demographics:
7. What is your gender?
☐ Male
☐ Female
8. What is your age?
9. Where do you live (City, State)
10. What is your profession?
11. Are you interested in participating in future tasks related to autism?

Please feel free to provide any additional comments you may have, we appreciate your input!

Thank you!

Figure 5.24 Screen shot of the qualification task for the autism topic. Asking for both demographic information and questions helped determine the expertise of the respondents with regard to the topic.

and to rate the apparent expertise of each response. Responses that were incomplete or obvious spam (e.g. answering “asdf” (the four left fingers on the keyboard) to several questions) were rejected (not paid). The topic experts were then asked to rate each response as it related to their topic on a “High-Medium-Low” scale. While conducting the ratings both topic experts decided that the initial “High-Medium-Low” scale was insufficient for classifying the responses in a way that was meaningful for their research, and they expanded the rating system to meet their needs.

In rating of expertise for the autism responses a fourth level of ‘None’ was added to the initial scale. In addition, a second binary rating (yes/no) was added to separate responses deemed useful regardless of ex-

Please explain your familiarity or knowledge of Autism Spectrum Disorders. (If you are not familiar Autism Spectrum Disorders please do not complete this survey.)

1. What is your gender?

- ☐ Male
☐ Female

2. What is your age?

3. Which of the following best describes your highest achieved education level?

4. What is your profession?

5. Where do you live (city, country)?

6. What is your nationality?

Lunchbox 1 (Matching Puzzle)

Matching the right images can open this lunchbox. The top of the lunchbox depicts an image. When completes the puzzle, the lunchbox opens and children can access their snack, cookie, lunch, or drinks.

Parents can adjust the images to the child's preferences. When a child learned the word, they can change it. You can relate images to the content of the lunchbox, such as 'bread' and 'milk', but this is not necessary.

The material is soft and the box has no cutlery, because they might hurt themselves with that.

The lunchbox fits the children, because it is easy to use and offers a learning process by means of text, images, and sounds. The image and sound will attract their attention and the color can be attuned to each child. It is innovative, because the lunchbox focuses on the child's learning process.



1.a. Do you think this lunchbox idea is appropriate for children with autism?

- ☐ Yes
☐ A Little
☐ No

1.b. Please explain, Why?

Figure 5.25 Screen shots
of the follow-up task for
the autism.

pertise. Similarly, for the topic of social connectedness the researcher also added a null category to rate expertise, but discarded the High-Medium-Low scale. Instead he classified the responses as providing “functional” or “experience” information and just like the autism ratings added a binary useful rating. The ratings were recorded for later comparison to future responses.

Using the MTurk web tools, all participants (regardless of expertise) were assigned a qualification and invited to participate in the second task for the same topic. The qualification not only granted them access, but can also be used by the participants as a filter to find tasks that they are qualified to accomplish. To further encourage participation, invitations were sent in the form of a \$0.05 bonus, along with a brief message thanking them for their previous response and requesting their partic-

ipation in the follow-up task. The follow-up task paid an additional 5¢ per response. After collecting pre-qualified responses for 14 days, the second set of tasks was opened to all workers, using the same procedure as the qualification task and requesting 150 responses.

These last two sets of responses were then given to the topic experts, to be rated using the same method used to rate the qualification responses. The analysis and comparison of the results of all three phases is presented in the results.

Results

Table 5.7 shows the number of responses received for each task. In total this study received 534 responses.

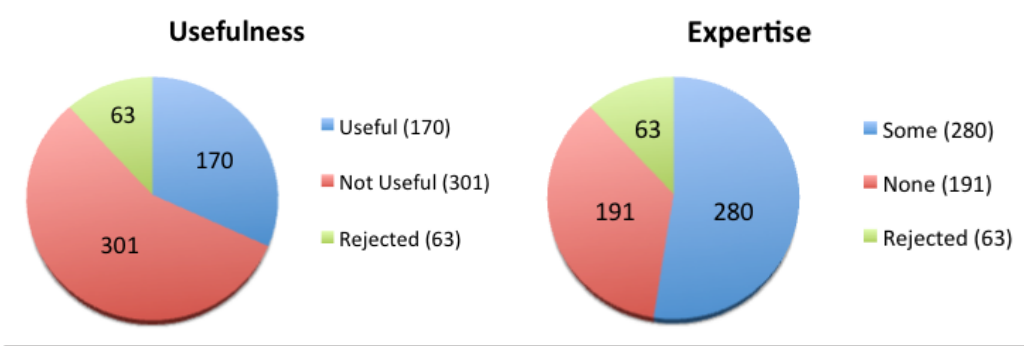
Table 5.7 The number of responses received for each task.

	Autism	Connectedness	Total
Task I: Qualification	123	168	291
Task II: Prequalified	11	9	20
Task II: Not Screened	71	152	223
Total	205	329	534

Figure 5.26 shows the distribution of the response ratings for both usefulness and expertise. To begin 63 (12%) were rejected as incomplete or spam responses, the remaining responses were rated on both expertise and usefulness. From the accepted responses, 280 (59%) demonstrated some level of expertise related to the topic, and 170 (36%) were identified as containing useful information. The 301 (64%) remaining responses were valid responses, but did not provide useful information according to the experts.

Figure 5.26 Overview of the usefulness and expertise ratings for all responses (n=534), as rated by the topic experts.

The data below takes a more detailed look at the response ratings for each study. The graphs show the breakdown of the expertise and usefulness ratings for each task and topic.



Qualification

The responses from the qualification tasks contained sufficient information for the topic experts to rate both the expertise and usefulness of the responses. Figures 5.27 and 5.28 show the distribution of ratings for both topics. They can both be divided into three categories: useful experts, non-useful experts, and throwaway responses. The researchers had not expected the middle group of responses, who demonstrate the requisite expertise, but did not provide responses that were useful or informative. While the distributions for topics were quite different, two similarities can be seen. First, there is the limited number of responses in the useful-expert category (the target users). This is especially true for the autism topic. The second trend is the high number of responses from India that are in the throwaway category, once again more prevalent in the autism topic, which has a more constrained target population.

Figure 5.27 Expertise and usefulness ratings of responses received from the qualification task, from the children with autism study (n=123). The 21 responses on the left demonstrate some level of expertise and provide information useful to the designer. On the right some responses indicated expertise, but did not include useful informations.

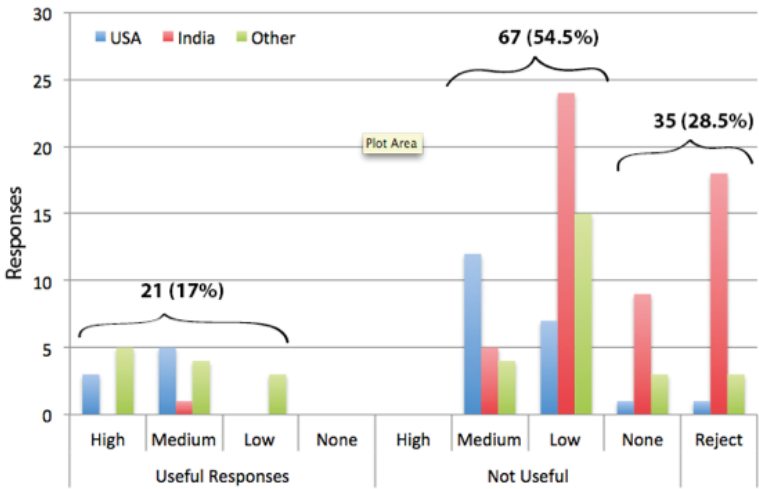
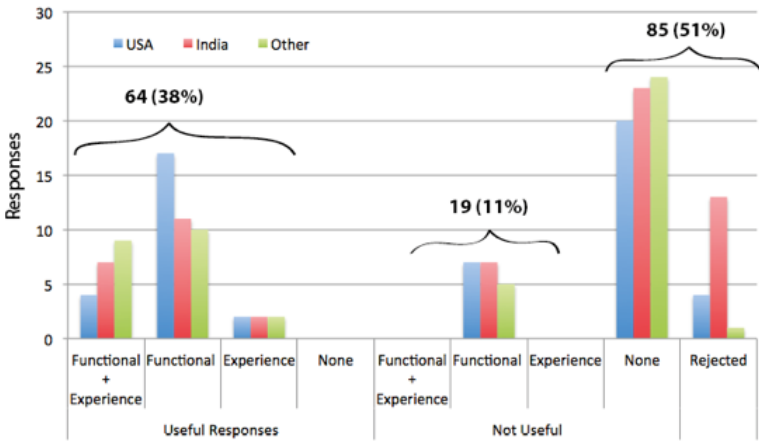


Figure 5.28 Expertise and usefulness ratings of responses received from the qualification tasks for the social connectedness study (n=166). There were significantly more responses that demonstrated the requisite knowledge and provided useful responses, and the poor responses were more easily identified.



Pre-Qualified Responses

The second phase of the study examined the effectiveness of a qualification task for filtering or identifying participants by comparing qualification ratings to ratings on a follow-up task. In spite of invitations and bonus payments, only a small percent (5% and 9%) of the original participants returned for the follow-up task. From the small sample there is evidence that the quality of responses remained relatively constant between the first and second task. As shown in Figure 5.29 the autism topic saw a small drop in the expertise ratings, while the connectedness topic shows a consistent rating with a small increase, Figure 5.30.

Figure 5.29 A comparison of the demonstrated expertise between qualification and follow-up task ratings for individual participants (N=11) who returned to complete the follow-up task for the autism study. The responses demonstrated a decrease in expertise on the follow-up task.

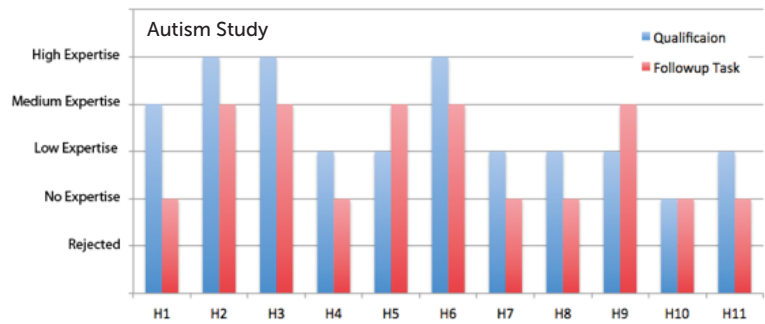
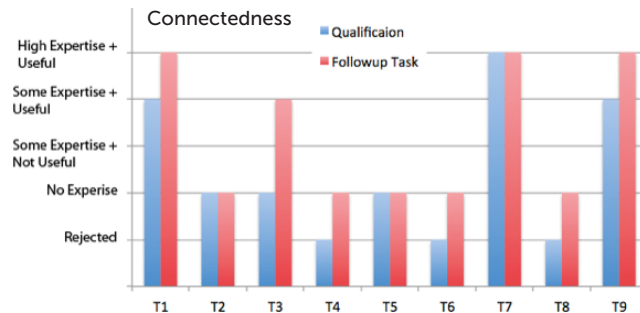


Figure 5.30 A comparison of demonstrated qualification and follow-up task ratings for individual participants (N=9) who returned to complete the follow-up task for the connectedness topic. The responses show stable or increased expertise on the follow-up task



Non-Qualified Responses

For the final phase of this study the same follow-up task was posted, but available to all workers on MTurk without the use of qualifications. For this phase 223 responses were received and once again rated by the topic experts.

The autism topic received less than half the number of responses compared to the qualification task. The ratings for both topics show similar distribution to the ratings from the qualification task. The two primary differences that can be observed are the increased polarization of the autism topic and less polarization of the connectedness topic.

Figure 5.31 Expertise and usefulness ratings for the second (without pre-qualification), autism task (n=71). In comparison to the prequalification responses (Figure 5.32) the distribution is relatively similar, with a low percentage of useful responses and an abundance of responses from India that are not useful.

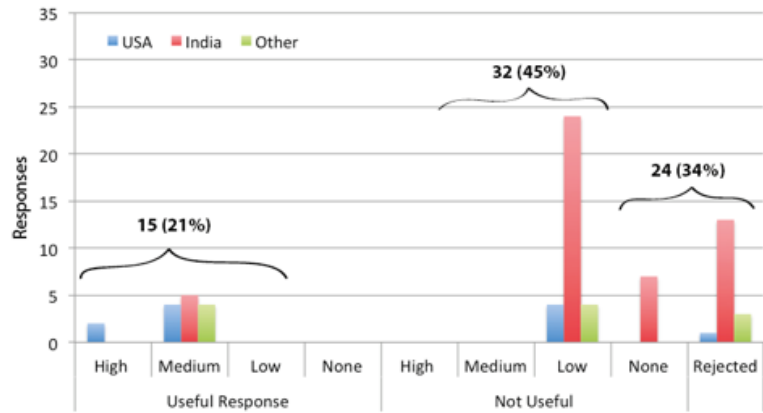
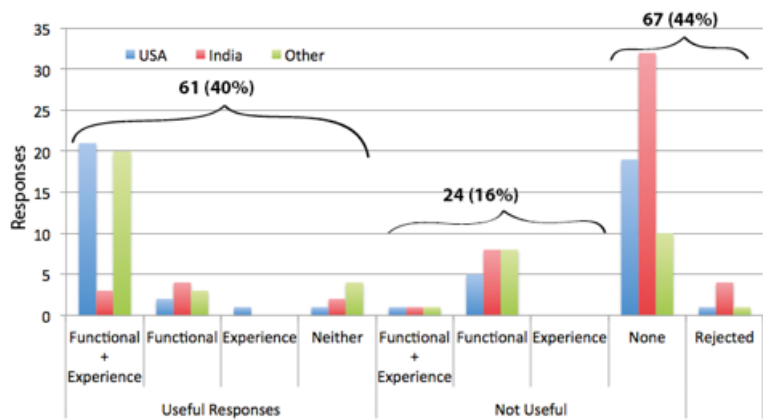


Figure 5.32 Expertise and usefulness ratings for the second (without pre-qualification), social connectedness task. In comparison to the prequalification responses (Figure 5.31) the distribution is more diverse, the combination of the qualification and information questions made finer distinctions possible.



These results show how the use of qualification tasks can identify and filter responses based on expertise, or identify members of a specific target user group. The information in the responses also makes it possible to filter based on quality. The following discussion further examines what these results mean in the context of crowdsourcing as a source of user information for design.

Discussion

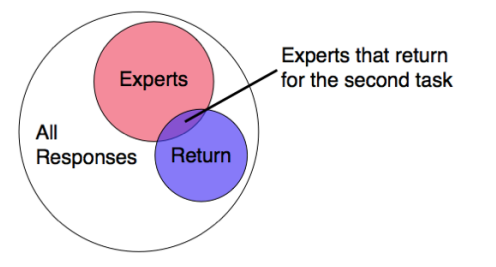
The results of this study demonstrate how qualification tasks can be used to identify respondents with the requisite expertise, as well as distinguish useful responses. The topic experts found it relatively easy to rate the responses, especially the exceptional responses (good or bad), but found it difficult to rate responses that fell in the middle. The middle of the road responses sparked considerable discussion among the researchers who wondered if they could be aggregated in some way to produce useful insight. In the end, it was agreed upon that time was better spent on the exceptionally informative responses and if additional

input or support was desired more results could easily be obtained by (re)posting a new task.

The second phase of the study shows that only a few people returned to complete a follow-up task (5% to 9%) despite invitations and bonus payments. This was likely a result of the delay, between completing the qualification and the availability of the second task (about one week), while the initial responses were collected and reviewed. Figure 5.33 shows how with separate layers of a tasks the number of available participants diminishes making it more difficult to reach them. A possible cause is that common qualification tasks on MTurk are used to train and filter workers for numerous follow-up tasks, giving the participant greater motivation to return. Our qualification tasks did not explicitly use this format.

Despite the small sample of returning participants, the quality (apparent expertise and usefulness) of each participant remained relatively constant between the first and second task. This consistency shows that a qualification task can function as an indicator of future performance, allowing researchers to gather a subset of participants that demonstrate a specific expertise. Unfortunately, the low return rate of qualified participants (8.9% and 5.4% respectively) limits the efficiency of crowdsourcing as a user research tool.

Figure 5.33 From the pool of all responses only a portion are experts. The number of original participants that return for a second task is also small. Where these two groups overlap leaves a limited group of qualified experts that will return for a given task (4.1% and 2.4% respectively), limiting the value of pre-qualification.



The third phase of this study demonstrates that target users can be identified directly from a user research task, without the use of qualification tasks. The topic experts rated the expertise and usefulness of responses for all tasks. This eliminated the need to design and rate responses to a separate qualification task and the need to entice participants to return. Open or discussion type questions were more reliable for qualifying expertise.

The distribution of response quality for both qualification and research tasks emphasizes a polarity (useful/useless) in the response quality. The

autism responses were more polarized, likely because people who have spent time with these children are not prone to common misconceptions. In contrast the social connectedness had a broader range of expertise, but the introspective questions revealed a polarity in who could communicate these thoughts effectively. The increased polarity in the second tasks can be attributed to the type of questions. The second autism task consisted of rating product designs. While it was possible to assess the expertise based on their explanations, it was far more binary. People either did or did not understand the needs of these children. In contrast, the second task for the social connectedness study asked questions that were personal and experiential. The answers provided a level of detail enabling the topic expert to make finer judgments on expertise and usefulness of the responses.

Conclusions for the Framework

This study showed how qualification tasks could be used to reach constrained or select target groups within crowdsourcing populations (Figure 5.34). And experts can be easily identified based on the qualities of their responses. However the low return rate of pre-qualified workers makes this step inefficient and impractical. The relative ease in identifying experts post-hoc all but eliminates the need to conduct pre-qualification tasks. Instead, incorporating qualification questions (open/descriptive) into the primary task is a fast and effective way to identify the desired expertise or target group from the qualities of their responses.

The type of qualification questions to use will depend on the topic and the desired expertise. These questions should not rely solely on demographic self-reported answers, but allow for open answers that can be assessed for expertise. This study found that questions allowing people to recount personal experiences worked well. In this way participants were able to demonstrate rather than report their relevant expertise. In addition to identifying expertise, the responses to the qualification questions often provided insights helpful for understanding the person, and providing additional topic information and context for their other inputs.

Looking back at our conceptual framework (Figure 4.2), the designers experience during this study showed it is more efficient to identify expertise at the end of the interaction with the users rather than during recruiting. The combination of qualification questions with information seeking tasks reduces the workload, by removing the preliminary task

Figure 5.34 The results of this study show that for user research tasks, it is best to combine qualification questions with the primary information-seeking task, and not as separate tasks (HITs). Analyses also demonstrated that expertise was relatively simple to identify during synthesis of the responses.



to pre-qualify participants. The combination also connects additional insights that increase contextual insights and helps with interpreting other responses. To enable this post-hoc identification, the task should incorporate questions that allow participants to reveal (not report) their expertise (e.g. What is your experience with ____? vs. Do you have experience with ____?).

Study 6: Tips for Design Practice

Tip 1: Embed qualification questions into the original task to identify target users or topic experts.

Tip 2: Do not expect Turkers to return for a second task.

Tip 3: Qualification tasks are more suited for training participants to improve quality on high volume "click work".

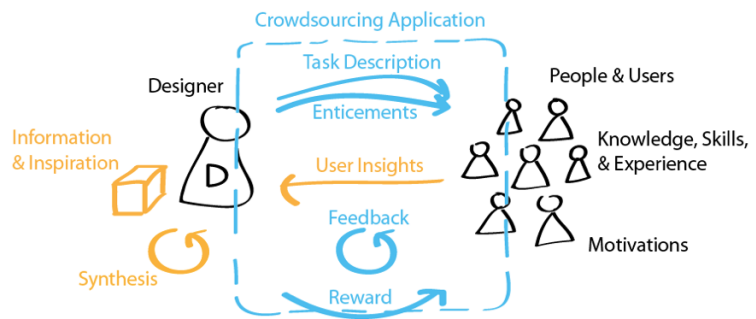
Tip 4: Combining qualification and information questions makes finer distinctions possible.

Tip 5: Know your topic well enough to identify misconceptions. They may provide wrong information if not filtered out, though they may prove useful in other ways.

CHAPTER 6 Discussion

The goal of this dissertation is to combine the benefits of online crowdsourcing applications with existing user research techniques, to provide designers with fast, abundant, and flexible access to users to inform the early stages of the design process. To that end, this research found success. Phase I explored how the crowdsourcing process operates when applied to design tasks. The results from phase I along with the literature review informed the development of the framework (Figure 4.2, reproduced below in Figure 6.1) to show designers the crowdsourcing process, aligned to their information needs. Phase II then explored the value of crowdsourcing for designers; demonstrating where designers can start, why the results are valuable, and how designers can engage users to support early design decisions.

Figure 6.1 The framework of crowdsourcing for designers as a fast, abundant, and flexible user research tool for early user centered design. The framework highlights the exchange between designers and users through a crowdsourcing application to elicit insights to inform and inspire the design process.



The following sections discuss both the research implications and the implications our results have for design practice. The final chapter presents the answers to the research questions and overarching conclusions along with a set of guidelines to support designers in conducting fast, abundant, and flexible user research. I also reflect on the limitations, issues, ethics, and directions for future research.

6.1 Research Implications

The research studies, as reported in Chapters 3 and 5 (phase I and II), reveal a number of findings and lessons for design that arise from the intersection of crowdsourcing and user research. The experiences gained through these studies identified benefits and barriers for designers. The fast, abundant, and flexible nature of crowdsourcing has shown to be a useful addition to the user research toolbox. The framework and guidelines developed through the studies depict the process and enable

designers to integrate the speed of crowdsourcing into their design process, access quality insights, and use qualifications to identify target users. To access users' insights through crowdsourcing involves more than the outsourcing of work typical in crowdsourcing. For user research, it is beneficial to view crowdsourcing as a means to enable communication and engage users, through the use of an online platform. As viewed by people familiar with crowdsourcing this is not a stretch, but for designers this relatively hands off approach to user research is a new idea.

The two studies in phase I were conducted to identify what happens inside the process and how crowdsourcing works for design tasks. This illuminates the role of the crowdsourcing application and helped to differentiate the results of crowdsourcing from that of a web search. The first study shows how the structure and process imposed by a crowdsourcing application and the social norms of the crowd influence the tasks and responses. The process of asking for users' insights instills a connection not present when searching the web. Designers more carefully consider what information they seek and responses are more consciously examined because they are responses to a specific request instead of the results of a general web search.

The second study, examining off-line crowdsourcing, highlighted the value of an online application in mediating communication with the crowd and the importance of participant motivations. The application's role in mediating communication is central to the speed and ease of crowdsourcing, freeing designers to focus effort on other activities. The participants' motivations in the off-line study supported previous research and emphasized the importance of intrinsic motivations (fun, sense of contribution, sharing of expertise) that drive self-selection and improve quality. Using these formative results the framework was expanded to more accurately represent crowdsourcing for user research.

The experience and findings from the studies in phase I combined with the literature reviews allowed for the development of the framework, capturing the process and communication links within the crowdsourcing process to help identify areas for further study. The first study in phase II employed an online logo design contest to illustrate the connections between the designers as hosts of the contest and each of the communication elements in the framework. The results of this study showed how the host's involvement in creating the task, providing feedback, and synthesizing the results brought insight into their design proj-

ect. The creation of the tasks forced the team to carefully consider the goal of their project (a key, yet easily overlooked step), to then clearly and concisely communicate it to the contest participants. During the contest, providing constructive feedback allowed the designers to think critically about the vision they had for their project. Although limited, the communication with the participant designers and the response of refined designs to feedback showed how other designers viewed and responded to the design goal to formulate their solution.

At the end of the logo contest, the syntheses of the results became far more interesting and informative than simply selecting the winning design. The large set of designs submitted by generally anonymous contributors (though some choose to disclose additional information to promote additional consulting services outside the crowdsourcing service) provided content as well as a context in which the team could candidly discuss the pros and cons of many designs. The designers as hosts found this experience liberating, as they were accustomed to having the designer present during design selection. In the end they did not use the winning design, but crowdsourcing provided much more than a winning solution. It provided both a process and content to help the team discuss and clarify its design goals. While the results are based on a single contest, they illustrate how crowdsourcing can contribute to the design process. By requiring the team to write a design brief, they had to formalize and make explicit their design goals. Engagement with contributors furthered their own design thinking, and they received an impressive variety of responses and ideas to spur their design thinking. Finally, the ideas provided the content and context to reconsider their goals and inspire their own design solution.

Study four embedded crowdsourcing as a user research tool into seven design projects. The results provided insight into where the framework fits into the design process. Incorporating crowdsourcing into several design projects reinforced the speed and flexibility of gathering a large set of responses from a diverse set of users. This new tool allowed the design teams to pre-test questions, add breadth or quantity to other results, reach users not readily available through other means, and begin gathering information early while preparations for other activities were still underway. Prior to receiving results, there was an initial hesitation from designers due to the change in mindset (getting inputs from anonymous people) and the inability to preselect participants (a key activity in most user research). Confidence grew as responses aligned with results

from familiar methods. In the end, design teams appreciated it as an addition (not a replacement) to their other user research activities. Because of its relative speed and simplicity for gathering responses, teams generally employed crowdsourcing in the early formative stages of their process, alongside Internet searches and initial site visits.

Synthesizing the large sets of responses presented a challenge for most teams. They were generally not accustomed to dealing with large data sets or responses that lacked the context gained from meeting in person or preselecting specific participants. Despite these struggles, researchers observed teams organically selecting rich responses and identifying promising design directions from the multitude of diverse responses. As depicted previously in Chapter 5 (Figure 5.12 in particular), their approach to synthesizing the responses changed. Rather than an in-depth exploration with a few select participants, individual insights were compiled to construct a holistic picture of the design space. While crowdsourcing often accomplishes work by parsing larger tasks into smaller pieces to be worked independently, user research does not benefit from the prearranged order of a deconstructed task. Instead, crowdsourcing responses are synthesized more organically. The organization of user insights was aided by incorporating open questions into the tasks. Though most responses were short and simple they still provided information, context, and indications of what users found most pertinent. Each set of responses included a couple of surprisingly detailed and thoughtful comments, providing additional depth and richness that helped bring context to other less detailed responses.

The fifth study focused on increasing the speed of responses by creating clear tasks that leveraged participant motivations. The results showed how thoughtful task design could generate response rates as high as 23 responses per hour, with as many as 10 responses in the first 10 minutes. While this does not match the speed of click-work, the increased speed over typical user research activities provided the opportunity to gather plenty of useful insights in the span of a three-hour workshop. This speed translates to nearly on-demand user responses. Although the tasks and responses in this study are simple, reducing the cycle time of typical user research would allow designers to gather user insights during the span of a design meeting, and conducting several simultaneous tasks would gather even more information in a short time. In contrast to results from literature that show a “sweet spot” relating reward to speed and quality, the research showed how emphasis on increasing speed did

not sacrifice quality. In our trials, the same considerations that increase speed (short, clear, simple, interesting) also support quality (more than 90% were judged usable by the design teams). More complex tasks or constrained expertise likely have additional trade-offs that warrant further study.

The final study found that the use of a qualification task for user research functioned quite differently than typical qualifications on MTurk. Usually, qualifications train participants for repetitive tasks and create a barrier to entry to increase quality. In user research, we used qualifications to seek experts. In this sense, the results showed that a separate qualification task was effective at identifying experts or target users, but only a limited few (<10%) returned to complete the primary task. Instead, we found that incorporating qualification questions into the primary task was effective and far more efficient for identifying useful responses and target expertise. For redundant tasks where users could be qualified for many similar tasks, the separation works well, but for complex or one-time tasks the combination was more appropriate and quite beneficial for adding context to their responses. Qualification questions that were open and gave users an opportunity to share a personal experience made it especially simple to determine expertise, provided priming for later questions, and added context for synthesis and analysis. While designers were still not able to select in advance who responded, the ease of distinguishing experts from among the responses allayed their concerns as they could easily discard (and not reward) inappropriate or irrelevant responses.

The next section focuses on how these results relate to design practice. In the final chapter the research questions are answered and the results are consolidated into a set of guidelines to help support designers incorporate crowdsourcing into their early user research activities.

6.2 Implications for Design

In conjunction with the research results are a number of implications for design practice to support the adoption of crowdsourcing as a user research tool. The framework (Figure 6.1) and guidelines (Table 7.2) are intended to enable designers to use crowdsourcing on their own to inform their design process. With the aid of these guidelines and the “how to” advice available on crowdsourcing websites it is reasonable to expect that designers can begin using this tool without further ado. As reported by the design teams in Study 5, crowdsourcing was an appreciated ad-

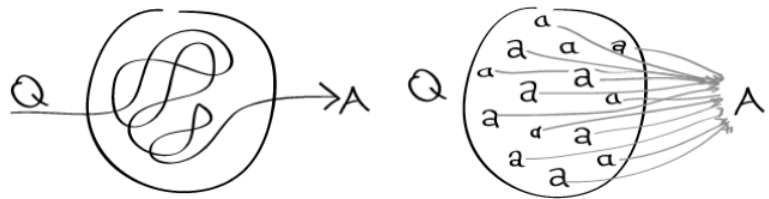
dition (though not a replacement) to their other user research activities. The speed and flexibility created new opportunities to gather early insights, pretest questions, and to triangulate with other results. Due to the reach of online tools, designers were also able to engage a much larger and more diverse user base. And while not all target groups are present in the crowd (e.g. toddlers), this expanded group of users can provide designers with expanded perspectives, and provides users new opportunities to be a part of the design process. Early on designers found it challenging to view crowdsourcing as much more than a web search, as it runs counter to current trends of empowering users by engaging them in participatory design. But with a little experience they began to see crowdsourcing as a complimentary means to engage and communicate with a large and diverse crowd of users. This new way of thinking opens up crowdsourcing as a user research tool, allowing crowdsourcing applications to be used to connect and communicate with users in fast, abundant, and flexible ways to inform the design process.

In this view of user research, crowdsourcing reveals some interesting implications for design practice. From the process, fit, speed, and ability to reach target users, each study reveals benefits that can be leveraged by designers. First, the imposed process of hosting a crowdsourcing activity encourages designers to pause and consider their goals and information needs, providing early clarity to the design team. Synthesis of a large set of crowdsourcing results not only begins to inform design decisions, but also helps in planning and focusing future, more involved, user research activities. While many designers consciously or unconsciously follow a process of defining the project goals and information needs, the imposed structure in crowdsourcing makes these steps explicit and can aid in training new designers or in spurring the team's thinking on a tough or stalled project.

The demonstrated flexibility and relative ease of crowdsourcing make it a nearly seamless fit throughout the design process, complementing a variety of design activities. It also caters to various information needs: providing initial insights, pre-testing questions, seeking inspiration, exploring a new context, and adding quantity or diversity to other information sources. Designers appreciated crowdsourcing's speed and adaptability as a welcome addition to their toolkit. Specifically, designers unanimously viewed it as an addition and by no means as a replacement for the visceral insights garnered through face-to-face interactions or in-situ observation. The online asynchronous nature of crowdsourcing

also provides uncommon access to a large and diverse population, adding quantity to confirm other responses and new viewpoints as contrast to the target user population. From the users perspective the reach of online tools allows more prospective users to have a voice and engage in the design process. While individual interactions may not be as visceral and engaging as face-to-face methods it certainly opens the door for more users to participate in the process. These added users and their insights provide a breadth of information that opens avenues to new insights.

Figure 6.2 visual depiction of gaining insights from users. The left diagram shows a question answered by traveling a single path that winds through the design space in order to reach the rich insights desired from a single user. In contrast, the diagram on the right depicts a crowdsourcing paradigm where richness can be gained by collecting numerous inputs from people throughout the design space and creating a composite understanding.



The speed of crowdsourcing fundamentally changes the accessibility of user inputs from weeks to hours. This transformation allows designers the opportunity to bring the user insights into the formative stages of design process. Leveraging the speed, designers can help ensure strategic decisions are based on insights from users without delaying the project. In addition, on-demand access to users allows questions that arise throughout the design process to be answered clearly rather than assumed or surmised.

The inability to pre-screen participants was a common concern among designers, as they are accustomed to spending considerable time finding, screening, and recruiting the right participants. To address the concern, the use of qualification tasks and questions was explored. While this does not allow for pre-screening, the integration of qualification questions into the primary task made it simple to identify target users and determine their levels of expertise from the responses. The ease in identifying useful responses may make filtering possible without adding questions, depending on the openness of the questions and complexity

of the topic or questions. The inclusion of open questions, especially when they allow users to describe an experience, was well suited as qualifications. Self-reporting was avoided, as it was insufficient for providing contextual information and difficult to verify. The addition of qualification questions increased task size for the user and increased synthesis of larger responses on the researcher's part. These open questions not only allowed for identifying target-users, but also provided additional information and often added contextual insights to help interpret responses to other questions. These insights helped bridge the digital divide between designers and the crowd.

For designers, crowdsourcing is not a replacement for the empathy and insights gained through in person interactions with users. In contrast, it provides a complementary tool with the ability to quickly reach a large and diverse population willing to provide user insights. Most importantly, the speed and flexibility of crowdsourcing provides access to bring user insights in near real-time to the strategic decisions made in the early formative stages of the design process, or in preparation for a field study, eliminating guesswork. Through increased experience, designers will refine their own approach and questions to access the nuanced insights they seek from users.

6.3 Limitations of Crowdsourcing User Research

Along with its many benefits, crowdsourcing has a number of limitations as a user research tool. These limitations can be generally associated with two root causes: the distant connection between designer and users via on-line application, and a disconnect between online asynchronous communication via crowdsourcing and traditional hands-on user research tools. Because of these limitations, designers view crowdsourcing as a useful addition to their user research efforts, but do not see it as a replacement.

One of the principle concerns from designers is the lack of face-to-face interaction with the contributors. Although we witnessed many diverse inputs being compiled to form rich insights and triangulation with other results, the lack of interaction or exposure to the users' context limits the visceral or nuanced connection to the users. The lack of context also made it difficult to interpret some responses. Interpretation was further hindered by the lack of nonverbal communication making synthesis difficult especially with the limited ability to ask clarifying questions. The feedback loop available in most crowdsourcing applications

provides a mechanism to send additional questions, but the time delay greatly limits responses to follow-up questions. This limitation becomes the key tradeoff between the speed and reach of crowdsourcing and the visceral, personal, and contextual insights gained interacting with users.

While, the elimination of participant recruiting is a primary component to the increased speed of crowdsourcing, designers felt that the lack of face-to-face interaction in recruiting participants inhibits empathy and relationship building – hallmarks of user centred design. This contrasts sharply with designers’ training and experience, which places significant emphasis on participant selection. The inability to control who participates is a natural concern for designers and can limit their commitment to the results. Adding to this are widespread misperceptions that only young men from India are participating in crowdsourcing. While participation from India is considerable, it is possible to explicitly include (Study 3), exclude (Study 4), or separate/filter responses from by country (Study 5).

Spam responses also inhibit crowdsourcing’s usefulness. Careful task design can reduce spam and make analysis easier, but some unusable responses will inevitably be received. The inherent simplicity and low cost of crowdsourcing make it easy to neglect parts of the process or not take the responses seriously. This was observed primarily in designing the task and in the synthesis of the results, limiting the value of the insights provided by users. The presence of spam can cause designers to mistrust the remaining responses. Most designers overcame their concerns through increased experience and corroboration of results from other user research, but this is a significant hurdle to overcome before adopting crowdsourcing as a user research tool. While crowdsourcing was appreciated for triangulation or for providing diversity, additional research is needed to further support designers.

Despite the limitations above, crowdsourcing provides a fast, abundant, and flexible tool for conducting user research. The speed and flexibility make it feasible to bring user insights into the formative stages of the design process. The final chapter summarizes the answers to the research questions, provides conclusions for the framework, and guidelines to support designers. It closes with a brief reflection on ethics and opportunities for future research.

Conclusions & Guidelines

Throughout the research journey, each study addressed a research question that built toward the main research question: How can designers combine the benefits of crowdsourcing with user research to bring information from users into the formative stages of the design process? With the primary aim of providing designers with a fast, abundant, and flexible means to gather users' insights to inform the design process. The research questions for each study were derived from the concerns or skepticism expressed by fellow designers and design researchers. In phase I, the first two studies looked at how crowdsourcing works for design tasks by distinguishing it from common online information sources. After development of a framework, phase II examined the value of crowdsourcing for designers and how they can reach the desired information from users. Below are the answers to each of the studies, followed by conclusions for the framework and a set of guidelines to support designers in their adoption of crowdsourcing.

What distinguishes crowdsourcing from web search as a source of user information for design?

The first of two studies in phase I focused on the differences in online sources of images to inform design. The results showed that each source produced a different set of images, where the perceived value, as a source of information, was influenced by the involvement (or lack) of the crowd. The reasons why images are available are an indicator of how the crowd is involved (search results, stock images, individual responses). While the crowdsourced set from MTurk was not the most preferred, the diversity and unexpected images triggered the most questions and discussion regarding the target users and their unexpected context or alternative viewpoints. While crowdsourcing requires additional time and effort (hours or days versus instantaneous), it is the process of requesting and receiving responses that connects the designer with users. It is the connection with users that distinguishes crowdsourcing from other online sources.

What distinguishes online crowdsourcing from questioning a known audience (or offline-crowd)?

This study examined the use of a physical crowd as a source of insights from users, providing contrast to online crowdsourcing as a means to access participants and looks at their motivations for participation. The results show that physical crowds operate under similar principles to online crowds. Each participant reported a mix of motivations. The most common were: interest, little effort, and helping out. In contrast to crowdsourcing, there was no explicit reward and some participants reported a sense of obligation (unintended) due to the classroom setting of the crowd. A positive difference was a complete lack of spam or low quality responses, potentially connected to their sense of obligation. A downside to the offline crowd was the lack of a mediating application to present the task and gather responses. In this study, paper requests were handed out in class and responses gathered via email, requiring considerably more effort than a crowdsourcing application. The final distinguishing characteristic was the near 50% response rate from a relatively small crowd (200), in contrast to a small response rate from a much larger crowd online (thousands).

These two studies in phase I helped illuminate the crowdsourcing process for design tasks. In combination with the literature review, they informed construction of the framework (Figure 6.1). Phase II consists of four studies that delve into the value of crowdsourcing for design and look at how designers can make the best use of this tool to inform their design process.

What can designers gain from hosting a design contest?

The first study in phase II examined an online design contest as a source of information and inspiration, in contrast to the typical production of a designed solution. The contest itself surprised the host designers with numerous, diverse, and high quality logo designs (46 original and 74 refined designs). But it was not the logos themselves that provided the value for the designers. Instead it was the interaction and discussion of ideas initiated by the process that was most valuable. The discussions occurred throughout the crowdsourcing process starting with the creation of the design brief (task description) that forced the requesting team to reach consensus on their design goals. Next, providing feedback on the logos and the exchanges with designers as they refined their submissions to better meet the design goals. Lastly, the selection of the winning design focused less on selecting a winner and more on how

different logos could be used and what they would convey about their organization. These interactions between designers inside their team and with designers in the crowd provided valuable information for their design process.

What is the value of crowdsourcing as a user research tool in design?

This study imbedded crowdsourcing into seven design projects, to examine the fit and value for informing the design process. Teams were free to incorporate crowdsourcing, as they deemed appropriate to meet their projects' information needs. The crowdsourcing activities provided useful and project specific information for all seven teams. The research results showed that the design teams appreciated crowdsourcing as an additional (not replacement) information source. This sentiment was in contrast to the visceral impact of interacting with users in person. Crowdsourcing was used throughout the design process, primarily to pretest questions, provide early insights, add breadth to the results from other user research activities, and to reach outside the immediately available user population for a point of contrast. The design teams were please with the speed, quality, and quantity of responses, but occasionally struggled with the analysis of numerous qualitative responses. The designers also recognized that the clarity and quality of their questions greatly influenced the quality and usefulness of the responses. These results showed that despite the lack of face-to-face contact, crowdsourcing was a fast, abundant, and flexible tool that could add valuable user insights throughout the design process.

How fast is crowdsourcing for user research tasks?

Crowdsourcing is faster than typical user research, but how fast? This study examined the speed of crowdsourcing user research, seeking near real-time responses similar to a web search. Through the use of clear questions, minimized effort indicators, and ample reward, we received response rates in excess of 20 responses an hour. Despite the emphasis on speed, the intentionally clear task design resulted in clear and usable responses. While not instantaneous, crowdsourcing can generate quality user responses during the span of a meeting or lunch break.

How can crowdsourcing reach people with specific expertise or knowledge?

This study examined how to ensure responses come from a specific or target set of users, despite asking for responses from a large and diverse

crowd. Through the use of qualification questions imbedded in the task, the results showed that target users participated and could clearly be identified. Qualification questions that were open and allowed users to describe experiences provided clear indication of their expertise and provided additional insights and context for their responses. As can be expected these more detailed and targeted tasks require more time to gather responses. And while not all user groups are represented online, user groups that are in the crowd can be reached and clearly identified.

7.1 Conclusions for the Framework

From the research results, it can be concluded that through experience designers overcame their initial scepticism of crowdsourcing and found it to be a welcome addition to their existing user research activities. The fast, abundant, and flexible nature of crowdsourcing provided an early means for designers to gather initial user inputs and pre-test questions. Later in the design process, it added value with its ability to quantify, triangulate, diversify, and gather user inputs in near real-time.

In contrast to typical user research tools, crowdsourcing asks designers to accept that user insights can be gained from an undefined crowd online who may provide sparse or incomplete information. This hands-off approach is a significant departure from the increased user involvement and interactions of participatory user research techniques, where considerable emphasis is placed on incorporating users directly into the design process. The conclusions below are consolidated into two sections: a table summary of what was learned about each element in the framework and a set of guidelines to support design practice.

The structure of the framework as introduced in Chapter 4 remained stable and was reinforced by the results of the studies. Through the course of the studies, we gained a deeper understanding of each element in the framework as well as insight into their connections throughout the process. In other words, the research shows how crowdsourcing connects the designer to a crowd of users through a structured interaction that can quickly provide information for the design process. To gain a sense of how the process connects designers to users, Table 7.1 provides a consolidation of the research findings, highlighting a more informed depiction of the role of each element. The descriptions capture the unique aspects of crowdsourcing user research. This approach fills a gap in the landscape of user research tools by providing access to user insights very early in the design process. Additionally, the speed,

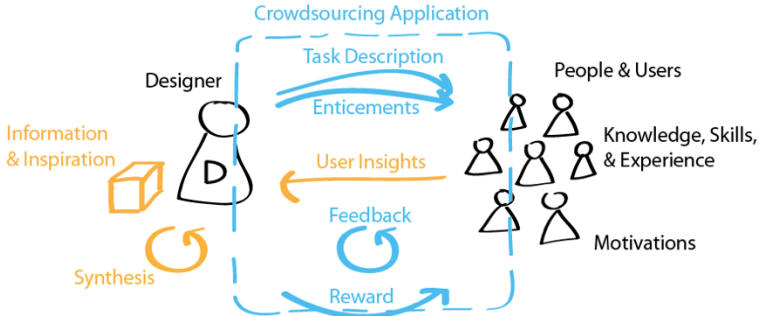
Table 7.1 A description of the elements of the framework, which includes insights gained through research with an emphasis on how they build connections to users in the crowd.

	Descriptions
Designer	The designer (or user researcher) is a central role and will select a platform, develop the task with enticements, receive the users' insights, provide feedback, pay the reward, and synthesize the insights into information and inspiration from the user, in order to meet their information needs.
Crowdsourcing Application	There are many applications and each offers a different environment (interfaces & options, typical tasks & rewards, size & makeup of crowd, etc.) for conducting user research. These differences influence the feasibility and appropriateness of a given application.
People & Users	The crowd is not a single entity, but a group of individual people, including members of our target user groups. Recognizing these people for their unique insights (no longer generic workers) is critical in asking for the insights they provide.
Knowledge, Skills, & Experience	To access specific user insights, the task (and enticements) needs to provide users with questions that encourage them to share and express their relevant insights while confirming their connection to the domain or topic of interest
Motivations	Understanding participant motivations helps designers develop the task description with enticements that activate participation and encourage quality inputs, from the desired user group, by balancing a mix of several motivational elements.
Task Description	The task description is the primary communication from designer to users. It needs to clearly describe what participants are asked to contribute while balancing enticements in order to gather the information desired from users.
Enticements	Enticements should be used to encourage appropriate and high quality participation from a specific group of users. Enticements should leverage intrinsic and explicit motivations to activate participation from users and discourage spam or low-quality responses.
User Insights	User insights are the primary communication from the users to the designer in response to the task. They contain the users' thoughts and ideas that help inform design decisions. Their quality and relevance will be tied to the selection of a suitable platform, the design of the task description, and the designers' ability to filter responses.
Feedback	The feedback mechanism built into most platforms provides a secondary means of communication between the designers and users. It is primarily used for praise, to explain rejected responses, or to encourage refinement of responses. The asynchronous interaction is a poor mechanism for follow-up or clarification of previous insights.
Rewards	The reward is the promised compensation for successful responses. It plays a significant role and can influence how and if users respond. Balancing reward with intrinsic motivations encourages users to provide the insights desired by user research.
Synthesis	Synthesis is an added level of analysis involved with interpreting user insights. This goes beyond the typical filtering and consolidation of responses. Synthesis is the process of both removing the irrelevant responses and translating the individual user insights into information and inspiration that can inform design decisions. This is where depth and richness of information can be built through aggregation of numerous diverse insights.
Information & Inspiration	The end result is synthesized information provided by the users in response to the designers' requests to inform and inspire the design process. This provides a deeper understanding of the users and their needs, resulting in more informed design decisions.

abundance, and flexibility provide new opportunities to pre-test questions, triangulate with other insights, and address questions as they arise throughout the design process.

Figure 7.1 (reproduced from 4.2, 6.1) visualizes the crowdsourcing process and the interplay between each of the elements as they relate to conducting user research. Through the evolution of the studies, specific insights were drawn to build a deeper understanding of each element and the relationships between them, tying together the elements in the framework. At the heart of the process, the designer as host controls many aspects of the process and gains insights from users. They also develop clarity of their goal by creating the task, build rapport with contributors by providing feedback, and see a detailed picture of the users' context through the synthesis of numerous small responses into a tapestry of the larger perspective. The crowdsourcing application is more than a facilitator of the process. The unique characteristics of each application influence the type of users participating and the types of responses they are accustomed (interested and willing) to answering. As we know from literature, the people participating are diverse and demographics shift between applications. Through research we found that incorporating qualification questions and appealing to motivations when designing the task descriptions can identify specific expertise or target groups. The studies also showed how simple well-designed task descriptions – that incorporated balanced enticements – could produce fast results with initial responses coming within minutes. The insights provided by the users' responses pleased researchers with their diversity, breadth, and depth, providing both information and inspiration for the design teams, despite the ever-present need to discard spam.

Figure 7.1 The final framework allowing designers to see the crowdsourcing process as a tool for user research. The framework also shows the interconnected nature of each element in the process tying together the descriptions in Table 7.1 to the guidelines in Table 7.2



Contribution to Research

The current studies explored the connection between crowdsourcing and user research as a tool for informing the design process. Through a series of six studies and the development of a framework, I examined and tested the connections, illuminating the possibilities and limitations. The conclusions above make the case for crowdsourcing as a new, fast, abundant, and flexible tool for user research. And through the development of the guidelines (below), I aim to make crowdsourcing accessible, enabling designers to connect with users early and often, as we pursue design solutions to the wicked problems of today and tomorrow.

7.2 Guidelines for Designers (and User Researchers)

These guidelines – based on the research – aim to enable designers to leverage the benefits of crowdsourcing into their own design process. The results demonstrate that the speed and flexibility of crowdsourcing can be successfully applied to bring user information into the design process faster, earlier, and more often. The results reveal that crowdsourcing is not a replacement for other more hands-on user research tools, nor is it ideally suited for all information needs. Despite overwhelming success, there were also frustrations among designers in the inability to pre-select who participates and how to handle irrelevant responses. Though these concerns were largely resolved through later studies and experience, designers still felt significant drawbacks to crowdsourcing. Accepting these limitations, designers used crowdsourcing as a flexible complement to other user research activities. Designers appreciated the speed and quantity of the results as initial insights, a means to pre-test questions for other activities, and as a way to triangulate or quantify other findings. In addition to the research findings and discussion, the guidelines in Table 7.2 (on next page) are presented as a step toward adopting crowdsourcing as a user research tool in your future design projects.

7.3 Recent Developments in Crowdsourcing

Since the completion of the data collection in 2012 there has been some relevant progress in crowdsourcing research. While not a comprehensive literature review, this brief section captures some key developments by a few leading researchers in crowdsourcing. In contrast to the enclosed research, recent developments are often focused on developing collaboration or cooperation among crowd workers in the pursuit of enabling complex tasks. Rather than informing the design process, they are building new custom platforms (or new interfaces for MTurk) that

Table 7.2 Guidelines to aid designers' implementation of crowdsourcing for user research..

	Guidelines
First, adopt a new perspective	This is not a deep dive or codesign process. Crowdsourcing is a fast means to access a large population. Tasks can be targeted and complex, but even simple responses can provide meaningful insights, especially when aggregated.
Use the help guide	Every crowdsourcing site has a guide of some type. Read it thoroughly and it will answer a lot of your questions and help you build tasks as intended by the site owners. Your remaining questions will be answered through experience.
Play nice	Ethics are a persistent concern in anonymous online communities. Be a valuable member to the community both in hosting and contributing. Know and follow the rules and norms, report violations, and encourage a positive experience for all.
No recruiting	Recruiting participants up front is not recommended. The elimination of recruiting is a significant contributor to the speed of crowdsourcing. And it is easier to select participants from their responses.
Be a contributor	By building experience with how people search for, select, and complete tasks you will learn to choose the right platform and design better tasks. Equally important, these are communities with social norms and responsibilities that are better learned as a contributor.
Delete the junk	Regardless of the platform and the quality of your task you will always receive some spam, generally less than 10%. It is easy to identify (e.g. 'asdfg') and remove before synthesis. If you get too much spam, revise and repost your task.
Early and often	The greatest benefit of crowdsourcing is speed. Use it to your advantage to bring user information into the process earlier. With near real-time responses you do not have to wait to find a user; post questions as they arise. Use later in the design process to add breadth, quantity, or triangulate.
Structure is your friend	The minimal but imposed structure of crowdsourcing can help your team clarify your project goals and information needs by writing them down. Synthesis of the diverse data sets can evoke candid discussion due to the disassociation with the contributors.
Reaching a target user group	Crowdsourcing operates on freedom of choice. Users choose if and when to contribute from the tasks available. Although not all user groups are represented in the crowd, if someone chooses your task, they likely believe they have something relevant to contribute.
Find expertise and target-users	Incorporate qualification questions into your primary task. Use open questions to ask about relevant experiences. The additional information makes identifying target-users easier and can provide useful contextual insights to interpret other responses.
Let the crowd surprise you	Include a couple open questions just for fun. Often they will be simple short answers of limited value, but occasionally someone will provide a long, interesting, deep, and insightful response. Send them a bonus!
Ask for pictures	The images users submit are largely from a web search, but they are selected by users (not an algorithm) in response to your question and can provide visual context. Pictures require a separate task when using the templates on MTurk.
Find the depth and richness	In contrast to a deep dive with a few users, a lot of individual responses from various users can be woven into a comprehensive picture of the design space, though it will lack the visceral impact of personal interaction. A card-sort is a nice place to start the synthesis of these large data sets. Do it again. If the responses you receive are not what you need or are incomplete, redesign your task and try again. The ability to ask follow-up or clarification questions is usually limited, so it is generally easier to resubmit a task than try to synthesize incomplete results.
Do not forget the basics	Remember you are conducting user-research. Use your fundamentals: be clear, know your goal, use demographics only as needed, use real Likert scales, ask reliable and valid questions, plan for post processing, etc.

enable the crowd to interact in ways that resemble the hierarchy and responsibility of traditional employment.

Michael Bernstein and colleagues at the Stanford Crowd Research Collective have developed a new crowdsourcing platform, DAEMO. Based on their previous research, DAEMO is touted as “easier and equitable crowdsourcing” giving contributors governing power over the platform. DAEMO builds on the lessons learned through the development and testing of Boomerang (Gaikwad et al. 2016), a unique reputation ratings system that realigns incentives and consequences for contributors and requesters. The ultimate goal of the reputation system is to increase the quality and complexity of crowd work by incorporating interactions and reputations that more closely mimic traditional employment. In addition, Kim, Agrawala and Bernstein (2017) developed Mosaic, a platform for the sharing of works-in-progress (graphic, video, books). In contrast to similar platforms where people share completed projects, Mosaic encourages sharing for the purpose of seeking feedback during the design process. This new form of crowd interaction has future implications for social computing platforms focusing on collaboration and skill development, by the crowd and for the crowd.

Steven Dow and a number of colleagues at Carnegie Mellon’s HCI Institute are exploring facilitation and structure in crowdsourcing as a means to improve work quality (Dow, Kulkarni, Klemmer & Hartmann 2012), crowd innovation (Chan, Dang & Dow 2016) and design critiques (Luther, Tolentino, Wei Wu, Pavel, Bailey, Agrawala, Hartmann & Dow 2015). Their results show that real-time feedback leads to better work, helps workers learn, and motivates more production.

Connecting the work of these two groups, Aniket “Niki” Kittur (2013) had previously proposed a framework for the future of crowd work. The framework sets a foundation to support a sustainable future for crowdsourcing, including the use of a hierarchy of workers to support more complex tasks. More recently he and his colleagues have been working on the Knowledge Accelerator (Hahn, Chang, Kim & Kittur 2016) to enable “crowd augmented cognition,” by connecting the strengths of machine learning and data mining with crowdsourcing and social computing.

As emphasized by these recent projects, the future of crowdsourcing research is focused on building collaboration in the crowd, ensuring qual-

ity of work, enabling more complex tasks, and integrating advances in social computing, all while balancing equity for crowd workers.

7.4 Issues & Ethical Concerns

The research studies were successful in receiving user insights and in answering the research questions. However, we did encounter issues related to methodology. And, while not a specific research area, we were alert to ethical concerns throughout this research. Below these concerns are discussed and we encourage mindful attention to fair treatment of participants when applying the above guidelines in your own crowdsourcing activities.

Methodology

The exploratory research approach provided a flexible architecture to try out and study this new user research tool, but this approach relies on a certain amount of serendipity to unearth valuable insights. Although the findings that were uncovered through this research were supported by the data and experiences gained during the research process, one wonders if a more structured methodology may have identified additional insights or provided stronger argumentation. Additionally, the exploratory research approach incorporated my active role as the researcher in many of the activities performed by the designers or participants in the studies. Although conscious attention was placed on the potential for bias, there still remains a possibility that the outcomes lean toward a more positive result than can typically be expected. Specifically, we cannot confirm if all spam responses were identified and removed (Was inaccurate information brought into the design process or were useful insights tossed out unnecessarily?). It is also possible we allowed confirmation bias to cloud our judgment regarding common but false perceptions regarding the context of the various design projects. While crowdsourcing demonstrated its value in a variety of settings, without further examination we should remain cautious of potential misperceptions that crowdsourcing may bring into the process.

This research also relied heavily on design projects conducted by students in an academic setting. The projects ranged from undergraduate workshops to graduate and doctoral research projects and, while several projects had corporate sponsors or students with professional design experience, they were conducted in the relatively controlled academic environment. In one case the entire crowd was composed of students. The academic environment advanced the research by providing structure

for the process and timeline of the design projects, but raised questions of transferability to industry. Fortunately the consistency of the results along with the diversity in topics, settings, and scale of these projects lends to their plausible transfer for use in industry.

Ethical Considerations in Crowdsourcing

While ethics was not a research topic, we were mindful of potential issues throughout this research. The anonymity provided by crowdsourcing presents unique ethical concerns, primarily in the fair treatment of the crowd. As an online application for exchanging work for money, there exists a possibility for abuse or exploitation. Fortunately the social nature and mediation of these applications provides a level of self-policing, where requesters and contributors alike shun and report individuals who do not conform to the rules or norms on a given site. This was witnessed first hand during the logo contest where the discussion section included an exchange where one designer felt his design was copied; the accused promptly removed their submission. Additionally there exist separate websites (e.g. turkopticon.ucsd.edu, turknation.com) where people can rate and leave feedback on requesters (hosts). The longevity of crowdsourcing applications is an indication that contributors are staying involved; in my experience contributors will quit, avoid specific types of tasks, or migrate to other crowdsourcing applications if they feel they are being exploited. Despite these layers of protection, there still exists the possibility of exploiting private information, copyright infringement, or of asking people to commit illegal or inappropriate acts such as click fraud (false product reviews or artificially increased web traffic) or CAPTCHA hacking (having people translate CAPTCHAs for use by automated systems). As designers, researchers, and now crowdsourcing requesters, we all have an obligation to treat contributors fairly and provide proper compensation for their efforts.

While concerns vary by activity and application, a common concern is the idea that workers are underpaid or treated as forced labor. This is a serious allegation, and a real possibility as the work performed can be more valuable than what workers are paid (e.g. idea competitions). Although this imbalance exists it discounts the non-monetary motivations that seem to dominate participants' decision to contribute. In addition, users have autonomy to select what, if any, tasks they contribute to, making it difficult for requestors to force participation. There does seem to be some evidence that in countries like India there are Turk 'shops' (ycombinator.com 2013) where requesters see abnormally high

response from a single location or IP address. The ‘shops’ reference implies that sweatshop type organizations exist. While there is a large active population of contributors from India, I did not personally encounter evidence of Turk-shops, possibly because our design tasks were not repetitive work that would enable numerous responses from a single person or rapid accumulation of rewards.

Another common concern is the exploitation of user information. In this regard, the applications encountered throughout this research all included participation agreements that expressly prohibit the collection of personal information (e.g. social security numbers, financial information). Contributors should watch out for and report requesters attempting to gather this information to administrators. These applications also have social constructs and behavioral norms that are enforced by the application, contributors, and hosts all help to mitigate abusive behavior.

Specific to design contests, professional designers have some legitimate concerns. The unauthorized use of submitted content (other than the paid winner) by the requester or even other contributors is a potential problem. While online design contests include participation agreements that spell out ownership of submitted content, they vary considerably and are hard to enforce. In general, it is difficult and expensive for an individual to fight copyright infringement. The other concern expressed by design professionals is the watering down or inclusion of amateurs in the design profession. While this may cut into job opportunities for trained and professional designers, there is no documented evidence that design contests are having either a positive or negative impact on the design profession. While not specifically examined in the current research, our experience suggests that crowdsourcing has opened a new market segment offering low cost options for small companies and organizations who can not afford professional design services. In this sense it may create work in the middle between DIY (do it yourself) design and professionals services. This middle ground appears to create opportunities for under employed, inexperienced designers, and design students to hone their skills or build their resume.

Many ethical concerns are generally alleviated by the structure of the crowdsourcing process, the active engagement, and social connection of participants. Still, all hosts and contributors have a responsibility to uphold moral behaviors and report suspected abuse, as we would during any other form of user research.

7.5 Future Research

The continuing growth and evolution of crowdsourcing applications creates a number of research opportunities that may provide additional benefit to designers, and possibly overcome some of the limitations identified above. First, the use of MTurk in these studies was rather rudimentary. In our interest to keep the tasks accessible for designers new to crowdsourcing, the tasks in this study relied on the provided templates with only minor formatting modifications, and did not take advantage of the application program interface (API) or other more intricate or collaborative crowdsourcing activities. Exploration into more complex modes of crowdsourcing that leverage the social and interactive participation could explore opportunities for collaboration and co-creation, in an effort to further bridge the digital gap between users and designers. Within crowdsourcing there exist opportunities to examine the use of more complex tasks in contrast to the reconstruction of many small responses from a larger and more diverse user set. More collaborative interactions may open the door to characterize and enhance the relationship building and higher levels of engagement that participatory methods seek. Additionally the similarities crowdsourcing has with open-development, and the ability to invite a much larger crowd to participate, may create opportunities to encourage appropriation and empowerment within the users population. In relation there is little know about the users' perspective in regards to engaging in the design process via a crowdsourcing application. We have seen their responses to questions and their willingness to provide insightful inputs, but what level or type of engagement is preferred by our increasingly connected users populations?

More fundamental research opportunities are present in examining other crowdsourcing applications, especially those with unique structures or hosted in languages other than English. Crowdsourcing in other languages could help target specific user groups, or explore how cultural differences influence the interactions between requester and contributors. These cultural differences were evident when one student group (in Study 4) replicated their MTurk task on Taskcn.com (a similar Chinese language site) and received more expressive responses to identical questions. Additional opportunities exist for a wide variety of comparative studies examining effort, speed, and quality among various crowdsourcing applications or in contrast to common user research methods. The value of non-user inputs for design or a study of the depth and richness of user insights and their impact on the resulting design could also ben-

efit the design and user research communities. There are also opportunities for dedicated work on the ethical concerns to ensure fair and equitable treatment of participants and their intellectual property.

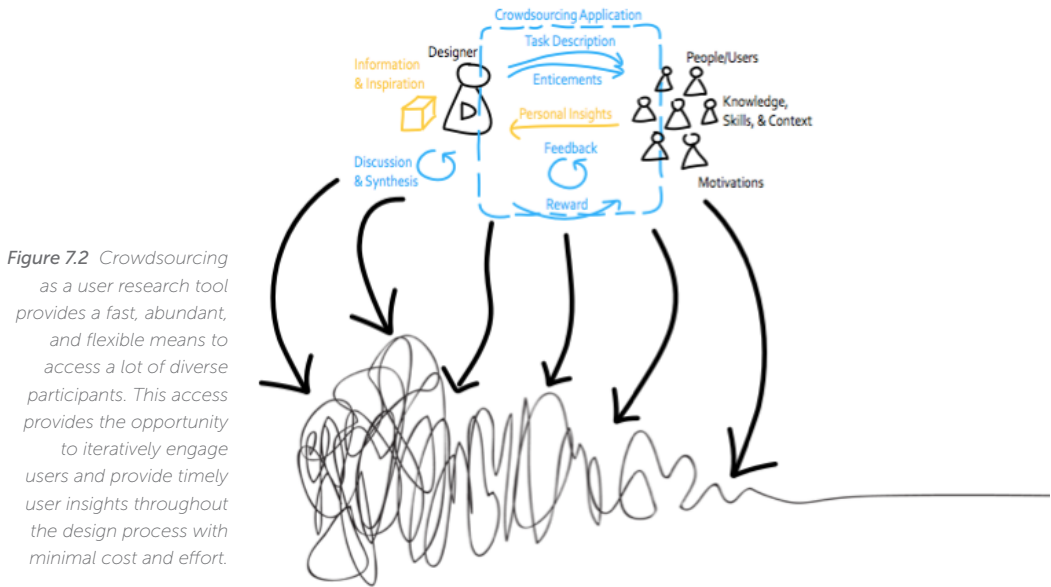


Figure 7.2 Crowdsourcing as a user research tool provides a fast, abundant, and flexible means to access a lot of diverse participants. This access provides the opportunity to iteratively engage users and provide timely user insights throughout the design process with minimal cost and effort.

It can be concluded that the presented framework provides a foundation for designers to branch out and apply crowdsourcing in their own way to gain access to the user insights throughout the design process (Figure 7.2). In addition, crowdsourcing can contribute to advancing user engagement in design practice, showing promising new opportunities for designers to access the user insights they need and desire.

References

- Alonso, O., Rose, D. and Stewart, B. (2008). Crowdsourcing for relevance evaluation. In *Proceedings of SIGIR Forum*, 42(2), November 2008, pp. 9-15.
- Archer, B. (1995). The Nature of Research. *CoDesign, interdisciplinary journal of design*, 2, pp. 6-13.
- Argyris, C. (1999). *On organizational learning, second edition*. Oxford, UK, Blackwell.
- Barbosa, L. and Feng, J. (2010). Robust sentiment detection on Twitter from biased and noisy data. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters (COLING '10)*. Association for Computational Linguistics, Stroudsburg, PA, pp. 36-44.
- Bernstein, M., Little, G., Miller, R., Hartmann, B., Ackerman, M., Karger D., Crowell, D. and Panovich, K. (2010). Soylent: a word processor with a crowd inside. In *Proceedings of the 23rd annual ACM symposium on User interface software and technology*, New York, NY.
- Beyer, H. and Holtzblatt K. (1997). *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann Publishers, San Francisco, CA, USA (1997).
- Boeijen, A., Daalhuizen, J., Zijlstra, J. and Schoor, R. (2014). *Delft Design Guide: Design strategies and methods*. BIS Publishers, Amsterdam, NL.
- Borst, I. (2010). *Understanding crowdsourcing: Effects of motivation and rewards on participation and performance in online voluntary activities*. PhD dissertation. Erasmus Research Institute of Management, Erasmus University, The Netherlands.
- Brabham, D. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, 14(1), pp. 75–90.
- Brabham, D. (2010). Moving the crowd at Threadless: Motivations for participation in a crowdsourcing application. *Information, Communication & Society*, 13(8), pp. 112-1145.
- Brabham, D. (2013). *Crowdsourcing*. MIT Press essential knowledge series, Cambridge, MA.
- Brydon-Miller, M., Greenwood, D. and Maguire, P. (2003). Why action research? *Action Research*, 1(1), pp. 9-28.
- Chan, J., Dang, S. and Dow, S. (2016). Improving Crowd Innovation with Expert Facilitation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, pp. 1223-1235.
- Chi, E. (2009). Augmented social cognition: using social web technology to enhance the ability of groups to remember, think, and reason. In *Proceedings of the 35th SIGMOD international conference on Management of data*. Providence, RI, pp. 973-984
- Chilton, L., Horton, J., Miller, R. and Azenkot, S. (2010). Task search in a human computation market. In *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '10)*. New York, NY, pp. 1-9

- Cooper, C. (2011). *Development of a Methodology to Support Design of Complex Aircraft Wings*. PhD Dissertation. Faculty of Aerospace Engineering, Delft University of Technology, The Netherlands.
- Corbin, J. and Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13(1), pp. 3-21.
- Curran, S., Feeney, K., Schaler, R. and Lewis, D. (2009). The management of crowdsourcing in business processes, In *Proceedings of IM '09. IFIP/IEEE International Symposium on Integrated Network Management-Workshops*, pp. 77-78.
- DiPalantino, D. and Vojnovic, M. (2009). Crowdsourcing and all-pay auctions. In *Proceedings of the 10th ACM Conference on Electronic Commerce (EC '09)*. ACM, New York, NY, pp.119-128.
- Dow, S., Gerber, E., & Wong, A. (2013, April). A pilot study of using crowds in the classroom. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris, France, pp. 227-236.
- Dow, S., Kulkarni, A., Klemmer, S. and Hartmann, B. (2012). Shepherd the crowd yields better work. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, pp. 1013-1022.
- Elo, S. and Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62, pp. 107-115.
- Evans, B. and Chi, E. (2010). An elaborated model of social search. *Information Processing & Management*, 46(6), pp. 656-678.
- Erickson, T. (2010). Geocentric Crowdsourcing and Smarter Cities: Enabling Urban Intelligence in Cities and Regions. In *Proceedings of 1st Ubiquitous Crowdsourcing Workshop at UbiComp*, 2010.
- Erickson, T. (2013). Social Computing. In: Soegaard, Mads and Dam, Rikke Friis (eds.). *The Encyclopedia of Human-Computer Interaction, 2nd Ed.* Aarhus, Denmark, The Interaction Design Foundation.
- Estelles-Arolas, E. and Gonzalez-Ladron-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189-200.
- Feng, D., Besana, S. and Zajac, R. (2009). Acquiring high quality non-expert knowledge from on-demand workforce. In *Proceedings of the 2009 Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources (People's Web '09)*. Association for Computational Linguistics, Stroudsburg, PA, pp. 51-56.
- Fitts, P. (1954). The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. *Journal of Experimental Psychology*, 47(0) pp. 381-391.
- Fletcher, A. (2006). *Do Consumers Want to Design Unique Products on the Internet? A Study of the Online Virtual Community of Threadless.com and their Attitudes to Mass Customisation, Mass Production and Collaborative Design*. Unpublished Bachelor's thesis, Nottingham Trent University, UK. Retrieved from <http://www.hiphipuk.co.uk/AdamFletcher-ThreadlessResearch.pdf>, Retrieved March 2010.

- Gaikwad, S., Morina, D., Ginzberg, A., Mullings, C., Goyal, S., Gamage, D., Diemert, C., Burton, M., Zhou, S., Whiting, M., Ziulkoski, K., Ballav, A., Gilbee, A., Niranga, S., Sehgal, V., Lin, J., Kristianto, L., Richmond-Fuller, A., Regino, J., Chhibber, N., Majeti, D., Sharma, S., Mananova, K., Dhakal, D., Dai, W., Puryanova, V., Sandeep, S., Chandrakanthan, V., Sarma, T., Matin, S., Nasser, Nistala, A., Stolzoff, A., Milland, K., Mathur, V., Vaish, R., and Bernstein, M. (2016). Boomerang: Rebounding the Consequences of Reputation Feedback on Crowdsourcing Platforms. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, pp. 625-637.
- Gaver, B., Dunne, T. and Pacenti, E. (1999). Cultural Probes. *Interactions*, 6(1), pp. 21-29.
- Glaser, B.G., and Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine, New York, NY.
- Gould, J. and Lewis, C. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM*, 28(3), pp. 300-311.
- Greenberg, M., Easterday, M. and Gerber, E. (2015). Critiki: A Scaffolded Approach to Gathering Design Feedback from Paid Crowdworkers. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition (C&C '15)*. ACM, New York, NY, pp. 235-244.
- Grefe, R. (2011) *What's the Harm In Crowdsourcing?* www.aiga.org/whats-the-harm-in-crowdsourcing/, Retrieved 12 July 2011.
- Hagen, P. and Robertson, T. (2009). Dissolving boundaries: social technologies and participation in design. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*. Melbourne, Australia, ACM.
- Hahn, N., Chang, J., Kim, J. E., and Kittur, A. (2016, May). The Knowledge Accelerator: Big picture thinking in small pieces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2258-2270). ACM.
- Haythornthwaite, C. (2009). Crowds and Communities: Light and Heavyweight Models of Peer Production, In *Proceedings of HICSS '09, 42nd Hawaii International Conference on System Sciences*. pp. 1-10.
- Heer, J. and Bostock, M. (2010). Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, pp. 203-212.
- Horton, J. and Chilton, L. (2010). The Labor Economics of Paid Crowdsourcing. In *Proceedings of the 11th ACM Conference on Electronic Commerce*. ACM, New York, NY, pp. 209-218.
- Imre, H. (2007). Comparison of three methodological approaches of design research. *Guidelines for a Decision Support Method Adapted to NPD Processes*.
- Hossmann, T., Legendre, F., Nomikos, G. and Spyropoulos, T. (2011). Stumbl: Using Facebook to collect rich datasets for opportunistic networking research. In *Proceedings of World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium*. IEEE Publishing, pp. 1-6
- Howe, J. (2006). The Rise of Crowdsourcing. *Wired*, 14(6), <http://www.wired.com/wired/archive/14.06/crowds.html>, Retrieved Feb 2010.

- Howe, J. (2008). *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*, Crown Publishing Group, New York, NY.
- Huberman, B., Romero, D. and Wu, F. (2009). Crowdsourcing, attention and productivity. *Journal of Information Science*, 35(6), pp. 758-765.
- Hutchinson et al. (2003). Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*, pp. 17-24.
- Ipeirotis, P. (2008). *Mechanical Turk: The demographics*. Blog post. <http://behind-the-enemy-lines.blogspot.com/2008/03/mechanical-turk-demographics.html>
- Ipeirotis, P. (2010). Demographics of Mechanical Turk, In *working papers CeDER-10-01*, Noy York University.
- Kapelner, A. and Chandler, D. (2010). Preventing Satisficing in Online Surveys. In *Proceedings of The World's First Conference on the Future of Distributed Work (CrowdConf 2010)*.
- Kaptein, M.C., Weisscher, A., Terken, J.M.B. and Nelissen, H. (2009). Online contextmapping; using the opportunities of Web 2.0 for the contextmapping procedure. *CoDesign, interdisciplinary journal of design*, 5(4), pp. 213-228.
- Kaufmann, N., Schulze, T. and Veit, D. (2011). More than fun and money. Worker Motivation in Crowdsourcing – A Study on Mechanical Turk. In *Proceedings AMCIS 2011*, 11, pp.1-11.
- Keller, I., Stappers, P.J. and Voegindeweij, S. (2004). Supporting informal collections of digital images: organizing browsing and sharing. In *Proceedings of Dutch Directions in HCI*. ACM Press, pp, 17.
- Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popović, Z. and Jaskolski, M. (2011). Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature structural & molecular biology*, 18(10), pp. 1175-1177.
- Kim, J., Agrawala, M. and Bernstein, M. (2016). Mosaic: Designing Online Creative Communities for Sharing Works-in-Progress. *arXiv preprint arXiv:1611.02666*.
- Kittur, A., Chi, E.H. and Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. In *Proceedings CHI 2008*, ACM Press, pp. 453-456.
- Kittur, A., Nickerson, J., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M. and Horton, J. (2013). The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13)*. ACM, New York, NY, USA, pp. 1301-1318.
- Kittur, A., Smus, B. and Kraut, R. E. (2011). CrowdForge: Crowdsourcing Complex Work. In *Human-Computer Interaction*. Human-Computer Interaction Institute, School of Computer Science, Carnegie Mellon University, pp.1801-1806.
- Kumar, V. (2012). *Design Methods: A structured approach for driving innovation in your organization*. Wiley, New York
- Kumar Whitney, P. (2003). Faster, Cheaper, Deeper User Research. *Design Management Journal*, 14(2), pp. 50-57.

- Lakhani, K. and Panetta, J. (2007). The Principles of Distributed Innovation. *Innovations: Technology, Governance, Globalization*, 2(3), pp. 97-112.
- Lakhani, K., Jeppesen, L., Lohse, P. and Panetta, J. (2008). *The value of openness in scientific problem solving*. Harvard Business School Working Paper No. 07-050.
- Leimester, J., Huber, M., Bretschneider, U. and Krcmar, H. (2009). Leveraging Crowdsourcing: Activation Supporting Components for IT-based Ideas Competitions. *Journal of Management Information Systems*, 26(1), pp.197-224.
- Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues*, 2(4), pp. 34-46.
- Lintott, C., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M., Nichol, R., Szalay, A., Andreescu, D., Murray, P. and Vandenberg, J. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 389(3), pp. 1179-1189.
- Little, G., Chilton, L., Goldman, M. and Miller, R. (2009). TurKit: tools for iterative tasks on mechanical Turk. In *Proceedings of the ACM SIGKDD, Workshop on Human Computation (HCOMP '09)*. ACM, New York, NY, pp. 29-39.
- Luther, K., Tolentino, J-L. Wei Wu, Pavel, A., Bailey, B., Agrawala, M., Hartmann, B. and Dow, S. (2015). Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, pp. 473-485.
- Martin, B., and Hanington, B. (2012). *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Beverly, MA: Rockport Publishers.
- Maxman, A. (2010). Driving Innovation: Ready, Set, Go! *Cell*, 140(2), pp. 171-173.
- Mulder, I. and Kort, J. (2008). Mixed Emotions, Mixed Methods: The Role of Emergent Technologies in Studying User Experience in Context. In: S. Hesse-Biber & P. Leavy (eds.), *Handbook of Emergent Methods in Social Research*. New York, Guilford Publications, pp. 601-612.
- Newman (2006). The Squiggle of Design. <http://v2.centralstory.com/about/squiggle/>, Retrieved 11 Feb 2010.
- Nelson, J. (1992). Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. ACM, New York, NY, pp. 373-380.
- Nielsen, J., and Sano, D. (1994). SunWeb: User interface design for Sun Microsystem's internal web. In *Proceedings of the 2nd World Wide Web Conference '94: Mosaic and the Web*, pp. 547-557.
- O'Reilly, T. (2005). *What is Web 2.0*. 30 September 2005. <http://www.oreilly.com/go/web2>, Retrieved 11 Feb 2010.
- Preece, J., Rogers, Y. and Sharp, H. (2011). *Interaction Design: Beyond Human-Computer Interaction (3rd Edition)*. John Wiley & Sons Ltd., New York, NY.
- Poetz, M. and Schreier, M. (2012). The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas? *JPROD INNOV MANAG 2012 PDMA*, pp. 245-256.

- Ross, J., Irani, L., Silberman, M., Zaldivar, A. and Tomlinson, B. (2010). Who are the crowdworkers?: shifting demographics in mechanical turk. In *Proceedings of the 28th of the international conference extended abstracts on human factors in computing systems*. ACM, New York, NY, pp. 2863-2872.
- Sanders, E.B.N. (2000). Generative Tools for CoDesigning. *Collaborative Design*. Scrivener, Ball and Woodcock (Eds.) Springer-Verlag London Limited.
- Sanders, E.B.-N. and Stappers, P.J. (2008). Co-creation and the new landscapes of design. *CoDesign, interdisciplinary journal of design*, 4(1), pp. 5-18.
- Schmidt, L.A. (2010). Crowdsourcing for Human Subjects Research. In *Proceedings of CrowdConf 2010*, San Francisco, CA.
- Schrage, M. (1991) *Shared Minds*. Random House, New York, NY.
- Schuler, D. and Namioka, A. Eds. (1993). *Participatory Design: Principles and Practices*. Lawrence Erlbaum, Hillsdale, NJ.
- Schwarz (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54(2), pp. 93-105.
- Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations*, Penguin Press.
- Sleeswijk Visser (2009). *Bringing the everyday life of people into design*. PhD dissertation. Industrial Design Engineering, Delft University of Technology, The Netherlands.
- Sleeswijk Visser, F., Stappers, P.J., van der Lugt, R. and Sanders, E.B.N. (2005). Contextmapping: Experiences from practice. *CoDesign, interdisciplinary journal of design*, 1(2), pp. 119-149.
- Sleeswijk Visser, F., van der Lugt, R. and Stappers, P.J. (2004). The Personal Cardset-a designer-centered tool for sharing insights from user studies. In *Proceedings of Second International Conference on Appliance Design*, pp.157-158.
- Snow, R., O'Connor, B., Jurafsky, D. and Ng, A.Y. (2008). Cheap and fast - but is it good? Evaluating non-expert annotations for natural language tasks. In *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics, pp. 254-263.
- Soukhoroukova A., Spann M. and Skiera, B. (2007). Creating and Evaluating New Product Ideas with Idea Markets. Paper presented at *AMA's 18th Annual Advanced Research Techniques Forum (ART Forum)*, Santé Fe, NM.
- Spool, J. and Schroeder, W. (2001). Test web sites: five users is nowhere near enough. In *Proceedings CHI2001 Extended Abstracts*. Seattle, WA, pp. 285-286.
- Stappers, P.J. (2007). Doing design as a part of doing research. In R Michel (Ed.), *Design research now: essays and selected projects*. Basel, Birkhauser, pp. 81-91.
- Sterling, B. (2006). http://blog.wired.com/sterling/index.blog?entry_id=1492720, Retrieved 23 May 2010.
- UPA (2012). http://www.upassoc.org/usability_resources/about_usability/what_is_ucd.html, Retrieved 7 March 2011.
- von Ahn, L., Maurer, B., McMillen, C., Abraham, D. and Blum, M. (2008). reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science*, pp 1465-1468.

von Hippel, E. and Katz R. (2002). Shifting Innovation to Users via Toolkits. *Management Sciences*, 48(7) pp. 821-833.

Wickens, C., Lee, J., Liu, Y. and Gordon-Becker, S. (2004). *Introduction to Human Factors Engineering (2nd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ.

Witla, P. (2009) Crowdsourcing and Its Application in Marketing Activities. *Contemporary Management Research*, 5, pp. 15-28.

Wordle on Flickr by daniel_iversen - Creative Commons - Attribution 2.0.

Yang, J., Adamic, L. and Ackerman, M. (2008). Crowdsourcing and knowledge sharing: strategic user behaviour on taskcn. In *Proceedings of the 9th ACM conference on Electronic commerce*. ACM, New York, NY, pp. 246-255.

Ycombinator (2009). Blog discussion on MTurk 'Shops.' Retrieved from <http://news.ycombinator.com/item?id=1010375>, Retrieved 17 November 2015.

Publications from this Work

Tidball, B., Mulder, I. and Stappers, P.J. (2017). Crowdsourcing As An On-demand User Research Tool. Submitted to *ACM Transactions on Computer-Human Interactions (TOCHI)*.

Tidball, B., Mulder, I. and Stappers, P.J. (2011). Online Design Contests: A Network of Inspiration for Designers. In *Proceedings of the 4th annual conference of the International Association of Societies of Design Research (IASDR2011)*, Delft NL, October 2011.

Tidball, B., Mulder, I. and Stappers, P.J. (2011). Inspiring Design: Exploring online user generated information. In *Proceedings of CHI Sparks 2011*, Arnhem NL, June 2011.

Tidball, B. and Stappers, P.J. (2011). Crowdsourcing Contextual User Insights. Position paper for the *ACM SIGCHI Workshop on Crowdsourcing and Human Computation*, Vancouver BC, May 2011.

Tidball, B., Stappers P.J. and Mulder I. (2010). Models, Collections, and Toolkits for HCI: What can we learn? In *Proceedings of the 24th BCS Conference on Human Computer Interaction (HCI2010)*. HCI-Educators Workshop. Dundee UK, September 2010.

About the Author



Brian Esley Tidball was born on June 18, 1975 in Newport Beach, California. He studied Industrial Engineering at the University of Washington, and commissioned into the United States Air Force upon graduation. After serving in various technical and leadership roles in the Air Force he completed masters degrees in Biomedical Engineering and Human Factors Engineering, with a thesis on perceptions of trust and emotions in avatar design.

In 2006 he became an instructor at the United States Air Force Academy, teaching and conducting research in aviation human factors and systems engineering. Outside of work he enjoys designing, building, food, travel, and spending time with his family.

Author's Publications

Cooper, C., Homan, J. and Tidball, B. (2017). Integrating Systems Engineering Students in Capstones; a multi-spectrum characterization of interdisciplinary capstones. Accepted for *Proceedings of the 15th Annual Conference on Systems Engineering Research*, Los Angeles CA, March 2017.

Tidball, B., Gibb, R., Andre, T. and Tossell, T. (2009). Undergraduate Human Factors Curriculum at The United States Air Force Academy. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting*, September 2009.

Bellenkes, A. and Tidball, B. (2008). Aircraft System Safety: Military and Civil Aeronautical Applications. *Ergonomics*, 51(7), pp. 1122-1124.

Tidball, B., Prabhala, S. and Gallimore, J. (2006). Making Faces: Exploring Perceptions of Personality Based on Emotional Expressions. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, September 2006.

Tidball, B. (2006). *Designing Computer Agents with Facial Personality to Improve Human-Machine Collaboration. Masters Thesis*. Department of Biomedical, Industrial and Human Factors Engineering, Wright State University, Dayton OH, March 2006.

Kancler, D., Gorman, M., Wesler, M., Bachmann, S., Stimson, D., Curtis, C., Randall, C. and Tidball, B. (2005). Application of Cognitive Task Analysis in User Requirements Definition and Prototype Design. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, September 2005.

Gorman, M., Anders, S., Donahoo, C., Quill, L., Jernigan, J., Goddard, M. and Tidball, B. (2003). *Core Automated Maintenance System (CAMS) at the Flightline: Nellis AFB Structured Study Test Report*. National Technical Information Service, July 2003.

Jacobs, J., Whitaker, R., Byler, D., Lemery, D. and Tidball, B. (2003). *Maintenance Mentor*. National Technical Information Service, September 2003.

Revels, A., Quill, L., Snead, A., Gorman, M., Jernigan J. and Tidball, B. (2002). *Wearable Computers in Air Force Maintenance Operations*. National Technical Information Service, December 2002.

Summary

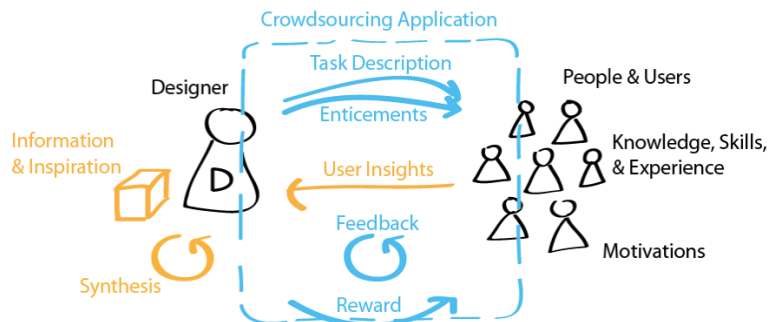
The design of products, services, software, and systems can benefit from information, insights, and inspiration from users. Numerous user research tools and methods have been developed to engage users and gather the desired inputs for the design process. Unfortunately, the time, effort, and expense of many tools often delay the design process or force designers to make decisions based on assumptions or incomplete information to keep the process moving. This dissertation investigates the use of existing crowdsourcing applications as a means to reduce the delay and inform early design decisions with end-user perspectives.

Crowdsourcing is one segment of the growing array of online activities where people contribute and interact. Crowdsourcing organizes some of these activities, providing access to a large online and on-demand workforce who voluntarily contribute to a wide variety of tasks (e.g. MTurk.com - click work, HatchWise.com - design competitions, SpeechPad.com - transcription). While numerous classifications exist, for this research crowdsourcing is constrained to activities where someone (in this case a designer or design researcher), uses an online platform to offer a reward to an undefined crowd (of potential users), for completion of a predefined task (providing user insights). This definition helps to distinguish it from similar activities like open-source development (e.g., Wiki, Linux), social networking (e.g. Facebook, Twitter), and crowd funding (e.g. Kickstarter, GoFundMe).

The current thesis investigates how designers can combine the benefits of crowdsourcing with user research techniques. The research consists of two phases. In phase I, a description of the crowdsourcing process was developed on the basis of literature review and two empirical studies. A framework placing the process into a design context was then developed, showing the key elements of the process and visualizing how crowdsourcing applications mediate a connection between designer and users. In phase II the framework is elaborated through a series of four studies. The results from phase II further inform the framework, show the value of crowdsourcing responses, and how to achieve desired results. The research culminates with a set of guidelines to make crowdsourcing more accessible for designers as a user research technique in the early stage of design.

The research was initiated in phase I with two formative studies (Chapter 3) where crowdsourcing was used to gather ethnographic insights. These studies examined the crowdsourcing process, identifying what happens and how it can be applied to user research. The results of the first study show that the connection with users (though not in person) differentiates crowdsourcing from other online sources of information, making crowdsourcing more interesting to designers. The second study concluded that off-line crowds are motivated by a mix of reasons, similar to crowdsourcing motivations: personal interest, low effort, and helping out. The second study also demonstrated the inherent value of the crowdsourcing application in mediating the interaction between designer and users for soliciting and gathering responses. The results of these two studies in combination with the literature review illuminated the key elements and interactions in the crowdsourcing process for user research and introduced its value as a user research tool.

A framework was initially formulated to illustrate the crowdsourcing process (Chapter 2). At the completion of phase I, the framework was updated to depict how a crowdsourcing application can operate as a user research tool connecting designer and users. The framework (below) depicts the process, showing the *crowdsourcing application* as the interface between *designer* and *users* (Chapter 4). The other elements depict the communication links, concluding with the *synthesis* of responses and the resulting *information*. With the framework in place, the four studies in phase II were conducted to further inform the framework and to address designers' questions regarding the value of crowdsourcing responses and how they can achieve the best results (Chapter 5).



The studies in phase II address research questions that were derived from the concerns or skepticism expressed by fellow designers and design researchers. The results (below) highlight the value and show designers how to best use crowdsourcing to inform their design process.

Logo Design Contest

In hosting a design contest, designers found the true value was not in the logos themselves, but in the interactions and discussion of ideas initiated by the crowdsourcing process. The development of a task description refined the teams design goals, providing feedback and engaging designers brought fresh ideas and further refined goals, and the selection of a winning design focused on how the design would be used and what it would convey about their organization.

Crowdsourcing in Design Projects

When crowdsourcing was imbedded into seven design projects, the designers appreciated the speed, quality, and quantity of responses. Crowdsourcing was viewed as a valuable and flexible addition (not a replacement) to their user research. It was used to pretest questions, provide early insights, add breadth or depth to the results from other user research activities, and to reach outside the immediately available user population. But the lack of face-to-face interaction detracted from the visceral connections expected in typical user research activities.

Speed for User Research

Through the use of clear questions, minimized effort indicators, and ample reward, we received response rates in excess of 20 responses an hour. Despite the emphasis on speed (not quality), the intentionally clear task design also resulted in clear and usable responses. While not instantaneous, the results generated quality user responses during the span of a meeting or lunch break.

Finding Target Users

By incorporating open-ended qualification questions into the task, target users could be easily identified within the crowd. While not all user groups are represented in the crowd, the open responses provided a clear indication of expertise and target users could be easily identified. The qualification responses also provided additional insights and contextual information to help interpret responses.

The results above indicated, that through experience, designers found crowdsourcing to be a complementary addition to their user centered design process. The fast and flexible nature of crowdsourcing provided an early means to gather initial user inputs and pre-test questions; later it was valued to quantify, triangulate, add diversity, and as a means to gather user inputs in near real-time. To leverage these advantages the results of the studies were consolidated into a set of guidelines.

The framework and accompanying guidelines (Chapter 7) capture the process of crowdsourcing as a fast abundant flexible tool for user research. This opens the door for designers to reach users in the beginning of the design process to inform early design process without making assumptions or delaying the process.

Samenvatting

Bij het ontwerp van producten, diensten, software en systemen maken ontwerpers gebruik van informatie, inzichten en inspiratie van en over gebruikers. Tal van gereedschappen ('tools') en methoden zijn ontwikkeld om gebruikers te betrekken om de gewenste input te verzamelen voor het ontwerpproces: gebruikersonderzoek ('user research'). Helaas vergen deze methoden vaak veel tijd, moeite en kosten, vertragen zij daarbij het ontwerpproces, of moeten ontwerpers beslissen op basis van veronderstellingen of onvolledige informatie om het proces in beweging te houden. Dit proefschrift onderzoekt het gebruik van bestaande *crowdsourcing* applicaties als een middel om de vertraging te beperken en vroege ontwerpbeslissingen te voorzien van perspectieven van eindgebruikers.

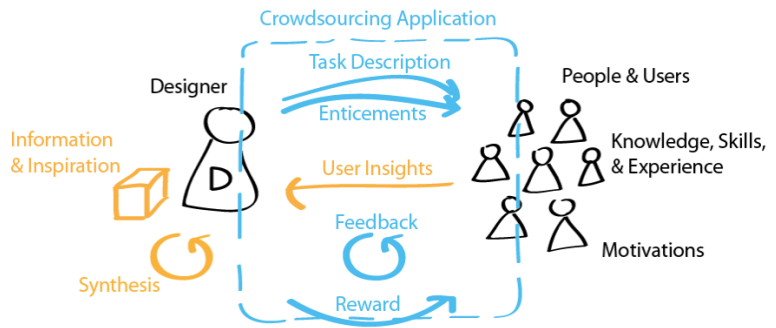
Crowdsourcing is deel van de groeiende reeks van online activiteiten waar deelnemers een bijdrage leveren aan een brede oproep. Crowdsourcing geeft opdrachtgevers toegang tot een grote verzameling van online en on-demand werknemers, die vrijwillig een bijdrage leveren aan een breed scala van taken (bijv. MTurk.com - 'clickwork', HatchWise.com - designcompetities, SpeechPad.com - transcriptie). Hoewel er tal van vormen bestaan beperkt dit onderzoek zich tot crowdsourcing als activiteiten waar iemand (in dit geval een ontwerper of ontwerp onderzoeker) gebruik maakt van een online platform om een beloning te bieden aan een 'crowd', d.w.z. een ongedefinieerde menigte (waaronder potentiële gebruikers), voor het uitvoeren van een vooraf gedefinieerde taak. In het bijzonder richt het onderzoek zich op taken die leiden tot inzichten over gebruikers. Deze definitie helpt om crowdsourcing te onderscheiden van soortgelijke activiteiten zoals open-source ontwikkeling (bijv., Wiki, Linux), sociale netwerken (zoals Facebook, Twitter) en crowdfunding (bijv. Kickstarter, GoFundMe).

Dit proefschrift onderzoekt hoe ontwerpers de voordelen van crowdsourcing met die van gebruikersonderzoek kunnen combineren. Het onderzoek bestaat uit twee fasen. In fase I is een beschrijving van het crowdsourcing proces ontwikkeld op basis van literatuurstudie en twee empirische studies. Vervolgens is een framework opgesteld dat het proces in een ontwerpcontext plaatst. Het framework identificeert de belangrijkste elementen van het crowdsourcing proces en visualiseert hoe

applicaties/platformen bemiddelen tussen ontwerper en gebruikers. In fase II wordt het raamwerk nader uitgewerkt in een serie van vier studies. De resultaten van deze fase helpen om het framework aan te scherpen, de waarde van crowdsourcing reacties te tonen, en leveren aanwijzingen om gerichte resultaten te verkrijgen. Het onderzoek wordt afgesloten met een set van richtlijnen voor ontwerpers om crowdsourcing toegankelijker te maken als onderzoekstechniek om gebruikers te betrekken in de vroege fase van het ontwerp.

Fase I van het onderzoek omvat twee formatieve studies (hoofdstuk 3) waarin crowdsourcing werd gebruikt om etnografische inzichten te verzamelen. Deze studies onderzochten het crowdsourcing proces, identificeerden wat er gebeurt en hoe de resultaten ervan kunnen worden toegepast voor gebruikersonderzoek. De resultaten van de eerste studie tonen aan dat het contact met gebruikers (al is het niet in levende lijve maar via een softwareplatform) crowdsourcing onderscheidt van andere online informatiebronnen. Juist dit contact maakt crowdsourcing interessant voor ontwerpers. De tweede studie concludeerde dat off-line crowds worden gemotiveerd door een mix van redenen, vergelijkbaar met crowdsourcing motivaties: persoonlijk belang, geringe inspanning, en de gelegenheid een vragensteller te helpen. De tweede studie toonde ook de intrinsieke waarde van van de crowdsourcing techniek om te mediëren in de interactie tussen ontwerper en de gebruikers bij het vergaren van inzichten. De resultaten van het literatuuronderzoek en deze twee studies werpen licht op de belangrijkste elementen en interacties in het crowdsourcing proces, toegepast voor het gebruikersonderzoek. Ook levert het zicht op de waarde ervan als een gebruiker-research gereedschap.

Een eerste raamwerk is ontwikkeld om het crowdsourcing proces te beschrijven (hoofdstuk 2). Bij de voltooiing van fase I is het raamwerk bijgewerkt om weer te geven hoe een crowdsourcing applicatie kan werken als onderzoeksinstrument dat designer en gebruikers in contact brengt. Het raamwerk (zie de illustratie hieronder) geeft de werkwijze aan om de applicatie in te zetten als interface tussen ontwerper en gebruikers (Hoofdstuk 4). De overige elementen tonen de communicatieverbindingen, afgesloten met de synthese van de reacties en de resulterende gegevens. Binnen dit kader zijn de vier studies in fase II uitgevoerd, deels om het kader verder te informeren, en deels om de twijfels en vragen van designers over de waarde van crowdsourcing resultaten te beantwoorden (hoofdstuk 5).



De onderzoeksvragen van de studies in fase II zijn afgeleid van de zorgen of scepsis van collega-ontwerpers en onderzoekers. De resultaten ervan (zie hieronder) verhelderen de waarde van de techniek, en geven richtlijnen voor ontwerpers hoe crowdsourcing het best ingezet kan worden voor het verkrijgen van gebruikersinzichten.

Logo design wedstrijd

Bij het uitschrijven van een prijsvraag, ervoeren ontwerpers dat de werkelijke waarde niet zozeer in het opgeleverde ontwerp (het winnende logo) zelf lag, maar vooral in de interacties en discussie van de ideeën tijdens het crowdsourcing proces. Het ontwikkelen van een taakomschrijving dwong de teams hun ontwerpdoelen te verfijnen. Het geven van feedback en het aangaan van ontwerpers bracht frisse ideeën en scherper zicht op de doelen van het ontwerp, en de selectie van een winnend ontwerp scherpste het zicht op hoe het ontwerp zou worden gebruikt en de boodschap die het zou uitstralen.

Crowdsourcing in design projecten

In zeven design projecten werd crowdsourcing ingezet. De deelnemende ontwerpers waardeerden vooral de snelheid, kwaliteit en kwantiteit van de reacties. Zij zagen crowdsourcing als een waardevolle en flexibele aanvulling van hun gebruikersonderzoek, maar niet als vervanging ervan. Crowdsourcing werd gebruikt om vragen uit te proberen en scherp te stellen, om vroegtijdig inzichten toe te voegen, om breedte en diepte toe te voegen aan de resultaten van andere onderzoeksactiviteiten, en deelnemers buiten de direct beschikbare participanten te betrekken. Maar het ontbreken van face-to-face interactie deed afbreuk aan de rijkheid van contact die ze van gebruikersonderzoek gewend waren.

Snelheid voor gebruikersonderzoek

Door gebruik te maken van duidelijke vragen, te letten op minimale inspanning en een ruime beloning voor respondenten, verkregen we

respons van meer dan 20 reacties per uur. Hoewel we de nadruk hadden gelegd op snelheid van de resultaten (niet kwaliteit), resulteerde het duidelijke taak ontwerp ook in duidelijke en bruikbare reacties. Hoewel het resultaat niet instantaan verschenen, betekent dit dat reacties van deelnemers van bruikbare kwaliteit niet langer duurde dan een reguliere vergadering of lunchpauze.

Het vinden van gebruikers binnen de doelgroep

Deelnemers uit de doelgroep bleken makkelijk te identificeren in de crowd middels gerichte vragen aan het eind van de interactie. Hoewel niet gegarandeerd kan worden dat alle gebruikersgroepen zijn vertegenwoordigd in de crowd, leveren de antwoorden op de open vraag een duidelijke indicatie van achtergrond van de gebruikers en of ze tot de doelgroep behoren. Deze antwoorden leveren ook extra inzichten op, en achtergrondinformatie om de resultaten van de taken te interpreteren.

De bovengenoemde resultaten geven aan dat ontwerpers crowdsourcing ervoeren als een waardevolle aanvulling voor hun user centered ontwerpproces. De snelle en flexibele aard van crowdsourcing maakt het tot een gereedschap om eerste gebruikersinput te verkrijgen en om de vragen voor veldonderzoek te verbeteren. Daarnaast werden de mogelijkheden herkend om kwantitatief te analyseren, resultaten van het veldonderzoek te trianguleren, om diversiteit toe te voeren. De bijna real-time snelheid van de resultaten blijkt ook een onderscheidende waarde. Op basis van deze studies werd een set van richtlijnen opgesteld om de techniek in de ontwerppraktijk toe te passen.

Het kader en de bijbehorende richtlijnen (hoofdstuk 7) specificeren het proces van crowdsourcing als een snel, rijk, en flexibel gereedschap voor gebruikersonderzoek. Dit opent de deur voor ontwerpers om gebruikers te bereiken en te betrekken in het begin van het ontwerpproces in plaats van enkel te bouwen op veronderstellingen of het proces uit te stellen.

Acknowledgements

This thesis was a long and interesting journey, and its successful completion would not have been possible without the support and contributions of many of you.

To my brilliant advisors, thank you. PJ your humble, caring, and innovative guidance makes you a true mentor. Your unwavering enthusiasm from our first contact and through out the intervening years provided inspiration and strategy to bring my ideas to life. Ingrid, it was great to work with you. You consistently challenged my perspectives, and forced me to delve into areas beyond my knowledge and experience to make this research and the writing the best it could be.

To the incredible people of the StudioLab: I truly appreciate the shared enthusiasm for all our respective projects, the intellectual sparring, and the multicultural perspectives. Walter, Jasper, Christine, Helma, Thomas, Daniel, Froukje, Nynke, Aadjan, Gert, Corrie, Caroline, Wei, Tal, Reinier, Stella, Elif, Natalia, Bia, Steven, Rob and many others, thank you all for an education way beyond my expectations and more importantly your friendship.

To my many mentors throughout my AF career: Gary Packard, Andy Bellenkes, Greg Ruderman, Dan Kennan, Randy Gibb, Terence Andre, Chad Tossell, Stu Turner, Aaron Byerley, Barb Masquelier, and Laurie Quill. Thank you for your trust, friendship, encouragement, and investment in making me a better leader, practitioner, and educator.

Sharon Bengel, your exceptional proof reading and editorial expertise transformed my jumble of run-on sentences into coherent text. Thank you!

Special thanks to the members of the Delft US military associated student society for their support and friendship! Cory Cooper, Greg Rickerd, Thomas DeNucci and your incredible families, I am truly grateful for the military comradery, shared adventures, holidays away from home, and your enduring friendship.

To my parents, Earl and Marcia: Thank you for a great start in life and the tools to learn from the world around me, the curiosity to try new things, and the drive to pursue my dreams.

I save my words of deepest gratitude for the limitless support of my favorite person, Cindy. Your encouragement, support, and patience made this possible. More importantly the care and dedication you give every day to raising our wonderful girls, I can't thank you enough. Laura and Natalie I love you dearly, and you no longer have to leave daddy alone while he writes.

Testimonial

While writing this dissertation a design team (one of its members had previously participated in Study 5) approached me to assist with crowdsourcing their preparatory research by pre-testing their questions before traveling overseas to conduct their primary user research efforts. The study was for a major beverage company who sought to tap into the hydration needs of yoga practitioners. After discussing the research findings and guidelines of this dissertation, we adjusted their questions and posted them on MTurk. Once they had a chance to look at the results the team sent the following email.

Hi Brian,

Thanks for sending us the results so quickly. I think the results are great. I didn't expect such fleshed out answers. Some of them are really surprising and insightful. Some of them really get into the problem we are trying to tackle with our projects which is reassuring! It is really interesting, that some of the people choose products specifically because of the enhancements and even add things to their liquids to make them more tailored to their needs. This is really the type of people we are aiming at. I'm definitely impressed that crowdsourcing can be used to get these type of results!!

Overall, this helps us moving forward as it highlights the general 'mindset' of yoga participants and that our target group of people at least put consideration into what they put into their bodies. It seems that this means that we can approach the group sessions with questions geared towards the nutritional needs from the product and types of enhancements people like.

If you are interested, I will keep you updated with the results of the research in the US and how much the findings corroborate with those from the crowdsourcing.

Thanks again for all your time and knowledge you have shared to get these results (and in helping us to formulate better questions in the future)! We really appreciate it!

*Best Regards,
G*



CROW



s t u
d i o
l a b

studiolab.ide.tudelft.nl